(51) International Patent Classification 5 : C12N 15/82, 15/32, 15/60 C12N 15/54, 5/10, A01H 5/00 A01N 63/02	**A1**	(11) International Publication Number: **WO 93/07278** (43) International Publication Date: 15 April 1993 (15.04.93)

(21) International Application Number: PCT/US92/08476

(22) International Filing Date: 5 October 1992 (05.10.92)

(30) Priority data:
772,027   4 October 1991 (04.10.91)   US
951,715   25 September 1992 (25.09.92)   US

(60) Parent Application or Grant
(63) Related by Continuation
US   951,715 (CON)
Filed on   25 September 1992 (25.09.92)

(71) Applicant *(for all designated States except US)*: CIBA-GEI-GY AG [CH/CH]; Klybeckstrasse 141, CH-4002 Basle (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : KOZIEL, Michael, G. [US/US]; 509 Carolyn Court, Cary, NC 27511 (US). DE-SAI, Nalini, M. [US/US]; 107 Silverwood Lane, Cary, NC 27511 (US). LEWIS, Kelly, S. [US/US]; 1508 Rutoni Drive, Hillsborough, NC 27278 (US). KRAMER, Vance, C. [US/US]; 608 Dana Court, Hillsborough, NC 27278 (US). WARREN, Gregory, W. [US/US]; 324 Bond Lake Drive, Cary, NC 27513 (US). EVOLA, Stephen, V. [US/US]; 1013 E. Beechmont Circle, Apex, NC 27502 (US).

CROSSLAND, Lyle, D. [US/US]; 108 Pinchot Lane, Chapel Hill, NC 27514 (US). WRIGHT, Martha, S. [US/US]; 106 Prince William Lane, Cary, NC 27511 (US). MERLIN, Ellis, J. [US/US]; 8605 Stonegate Drive, Raleigh, NC 27615 (US). LAUNIS, Karen, L. [US/US]; 104 Saddleview Lane, Franklinton, NC 27525 (US). ROTHSTEIN, Steven, J. [US/US]; 15 Fieldstone Road, Guelph, Ontario N1A 1A5 (US).

(74) Agent: FOLEY, Shawn, P.; Ciba-Geigy AG, P.O. Box 12257, Research Triangle Park, NC 27709 (US).

(81) Designated States: AU, BB, BG, BR, CA, CS, FI, HU, JP, KP, KR, LK, MG, MN, MW, NO, PL, RO, RU, SD, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: SYNTHETIC DNA SEQUENCE HAVING ENHANCED INSECTICIDAL ACTIVITY IN MAIZE

(57) Abstract

DNA sequences optimized for expression in plants are disclosed. The DNA sequences preferably encode for an insecticidal polypeptides, particularly insecticidal proteins from *Bacillus thuringiensis*. Plant promoters, particular tissue-specific and tissue-preferred promoters are also provided. Additionally disclosed are transformation vectors comprising said DNA sequences. The transformation vectors demonstrate high levels of insecticidal activity when transformed into maize.

SYNTHETIC DNA SEQUENCE HAVING ENHANCED

INSECTICIDAL ACTIVITY IN MAIZE

This application is a continuation in part application of U.S. serial no. 772,027 filed October 4, 1991, which disclosure is herein incorporated in its entirety.

Field of the Invention

The present invention relates to DNA sequences encoding insecticidal proteins, and expression of these sequences in plants.

Background of the Invention

Expression of the insecticidal protein (IP) genes derived from Bacillus thuringiensis (Bt) in plants has proven extremely difficult. Attempts have been made to express chimeric promoter/Bt IP gene combinations in plants. Typically, only low levels of protein have been obtained in transgenic plants. See, for example, Vaeck et al., Nature 328:33-37, 1987; Barton et al., Plant Physiol. 85:1103-1109, 1987; Fischoff et al., Bio/Technology 5:807-813, 1987.

One postulated explanation for the cause of low expression is that fortuitious transcription processing sites produce aberrant forms of Bt IP mRNA transcript. These aberrantly processed transcripts are non-functional in a plant, in terms of producing an insecticidal protein. Possible processing sites include polyadenylation sites, intron splicing sites, transcriptional termination signals and transport

1

signals. Most genes do not contain sites that will deleteriously affect gene expression in that gene's normal host organism. However, the fortuitous occurrence of such processing sites in a coding region might complicate the expression of that gene in transgenic hosts. For example, the coding region for the Bt insecticidal crystal protein gene derived from <u>Bacillus thuringiensis</u> strain <u>kurstaki</u> (GENBANK BTHKURHD, accession M15271, <u>B. thuringiensis</u> var. <u>kurstaki</u>, HD-1; Geiser et al. <u>Gene</u> 48:109-118 (1986)) as derived directly from <u>Bacillus thuringiensis</u>, might contain sites which prevent this gene from being properly processed in plants.

Further difficulties exist when attempting to express <u>Bacillus thuringiensis</u> protein in an organism such as a plant. It has been discovered that the codon usage of a native Bt IP gene is significantly different from that which is typical of a plant gene. In particular, the codon usage of a native Bt IP gene is very different from that of a maize gene. As a result, the mRNA from this gene may not be efficiently utilized. Codon usage might influence the expression of genes at the level of translation or transcription or mRNA processing. To optimize an insecticidal gene for expression in plants, attempts have been made to alter the gene to resemble, as much as possible, genes naturally contained within the host plant to be transformed.

Adang et al., EP 0359472 (1990), relates to a synthetic <u>Bacillus thuringiensis tenebrionis</u> (Btt) gene which is 85% homologous to the native Btt gene and which is designed to have

2

an A+T content approximating that found in plants in general.
Table 1 of Adang et al. show the codon sequence of a synthetic
Btt gene which was made to resemble more closely the normal
codon distribution of dicot genes. Adang et al. state that a
synthetic gene coding for IP can be optimized for enhanced
expression in monocot plants through similar methods,
presenting the frequency of codon usage of highly expressed
monocot proteins in Table 1. At page 9, Adang et al. state
that the synthetic Btt gene is designed to have an A+T content
of 55% (and, by implication, a G+C content of 45%). At page
20, Adang et al. disclose that the synthetic gene is designed
by altering individual amino acid codons in the native Bt gene
to reflect the overall distribution of codons preferred by
dicot genes for each amino acid within the coding region of the
gene. Adang et al. further state that only some of the native
Btt gene codons will be replaced by the most preferred plant
codon for each amino acid, such that the overall distribution
of codons used in dicot proteins is preserved.

        Fischhoff et al., EP 0 385 962 (1990), relates to plant
genes encoding the crystal protein toxin of <u>Bacillus</u>
<u>thuringiensis</u>. At table V, Fischhoff et al. disclose percent
usages for codons for each amino acid. At page 8, Fischoff et
al. suggest modifying a native Bt gene by removal of putative
polyadenylation signals and ATTTA sequences. Fischoff et al.
further suggest scanning the native Bt gene sequence for
regions with greater than four consecutive adenine or thymine
nucleotides to identify putative plant polyadenylation signals.

3

Fischoff et al. state that the nucleotide sequence should be altered if more than one putative polyadenylation signal is identified within ten nucleotides of each other. At page 9, Fischoff et al. state that efforts should be made to select codons to preferably adjust the G+C content to about 50%.

Perlak et al., PNAS USA, 88:3324-3328 (1991), relates to modified coding sequences of the Bacillus thuringiensis cryIA(b) gene, similar to those shown in Fischoff et al. As shown in table 1 at page 3325, the partially modified cryIA(b) gene of Perlak et al. is approximately 96% homologous to the native cryIA(b) gene (1681 of 1743 nucleotides), with a G+C content of 41%, number of plant polyadenylation signal sequences (PPSS) reduced from 18 to 7 and number of ATTTA sequences reduced from 13 to 7. The fully modified cryIA(b) gene of Perlak et al. is disclosed to be fully synthetic (page 3325, column 1). This gene is approximately 79% homologous to the native cryIA(b) gene (1455 of 1845 nucleotides), with a G+C content of 49%, number of plant polyadenylation signal sequences (PPSS) reduced to 1 and all ATTTA sequences removed.

Barton et al., EP 0431 829 (1991), relates to the expression of insecticidal toxins in plants. At column 10, Barton et al. describe the construction of a synthetic AaIT insect toxin gene encoding a scorpion toxin using the most preferred codon for each amino acid according to the chart shown in Figure 1 of the document.

Summary of the Invention

The present invention is drawn to methods for enhancing

4

expression of heterologous genes in plant cells. Generally, a
gene or coding region of interest is constructed to provide a
plant specific preferred codon sequence. In this manner, codon
usage for a particular protein is altered to increase
expression in a particular plant. Such plant optimized coding
sequences can be operably linked to promoters capable of
directing expression of the coding seqence in a plant cell.

        Specifically, it is one of the objects of the present
invention to provide synthetic insecticidal protein genes which
have been optimized for expression in plants.

        It is another object of the present invention to
provide synthetic Bt insecticidal protein genes to maximize the
expression of Bt proteins in a plant, preferably in a maize
plant. It is one feature of the present invention that a
synthetic Bt IP gene is constructed using the most preferred
maize codons, except for alterations necessary to provide
ligation sites for construction of the full synthetic
gene.

        According to the above objects, we have synthesized Bt
insecticidal crystal protein genes in which the codon usage has
been altered in order to increase expression in plants,
particularly maize. However, rather than alter the codon usage
to resemble a maize gene in terms of overall codon
distribution, we have optimized the codon usage by using the
codons which are most preferred in maize (maize preferred
codons) in the synthesis of the synthetic gene. The optimized
maize preferred codon usage is effective for expression of high

5

levels of the Bt insecticidal protein. This might be the result of maximizing the amount of Bt insecticidal protein translated from a given population of messenger RNAs. The synthesis of a Bt IP gene using maize preferred codons also tends to eliminate fortuitous processing sites that might occur in the native coding sequence. The expression of this synthetic gene is significantly higher in maize cells than that of the native IP Bt gene.

Preferred synthetic, maize optimized DNA sequences of the present invention derive from the protein encoded by the cryIA(b) gene in <u>Bacillus thuringiensis</u> var. kurstaki, HD-1; Geiser et al., <u>Gene</u>, 48:109-118 (1986) or the cryIB gene (AKA Crya4 gene) described by Brizzard and Whiteley, <u>Nuc. Acids. Res.</u>, 16:2723 (1988). The DNA sequence of the native kurstaki HD-1 cryIA(b) gene is shown as Sequence 1. These proteins are active against various lepidopteran insects, including <u>Ostrinia nubilalis</u>, the European Corn Borer.

While the present invention has been exemplified by the synthesis of maize optimized Bt protein genes, it is recognized that the method can be utilized to optimize expression of any protein in plants.

The instant optimized genes can be fused with a variety of promoters, including constitutive, inducible, temporally regulated, developmentally regulated, tissue-preferred and tissue-specific promoters to prepare recombinant DNA molecules, i.e., chimeric genes. The maize optimized gene (coding sequence) provides substantially higher levels of expression in

6

a transformed plant, when compared with a non-maize optimized gene. Accordingly, plants resistant to Coleopteran or Lepidopteran pests, such as European corn borer and sugarcane borer, can be produced.

It is another object of the present invention to provide tissue-preferred and tissue-specific promoters which drive the expression of an operatively associated structural gene of interest in a specific part or parts of a plant to the substantial exclusion of other parts.

It is another object of the present invention to provide pith-preferred promoters. By "pith-preferred," it is intended that the promoter is capable of directing the expression of an operatively associated structural gene in greater abundance in the pith of a plant than in the roots, outer sheath, and brace roots, and with substantially no expression in seed.

It is yet another object of this invention to provide pollen-specific promoters. By "pollen-specific," it is intended that the promoter is capable of directing the expression of an operatively associated structural gene of interest substantially exclusively in the pollen of a plant, with negligible expression in any other plant part. By "negligible," it is meant functionally insignificant.

It is yet another object of the present invention to provide recombinant DNA molecules comprising a tissue-preferred promoter or tissue-specific promoter operably associated or linked to a structural gene of interest, particularly a

7

structural gene encoding an insecticidal protein, and
expression of the recombinant molecule in a plant.

It is a further object of the present invention to
provide transgenic plants which express at least one structural
gene of interest operatively in a tissue-preferred or
tissue-specific expression pattern.

In one specific embodiment of the invention disclosed
and claimed herein, the tissue-preferred or tissue-specific
promoter is operably linked to a structural gene encoding an
insecticidal protein, and a plant is stably transformed with at
least one such recombinant molecule. The resultant plant will
be resistant to particular insects which feed on those parts of
the plant in which the gene(s) is(are) expressed. Preferred
structural genes encode B.t. insecticidal proteins. More
preferred are maize optimized B.t. IP genes.

Brief Description of the Figures

Fig. 1 is a comparison of the full-length native Bt
cryIA(b) gene [BTHKURHD], a full-length synthetic maize
optimized Bt cryIA(b) gene [flsynbt.fin] and a truncated
synthetic maize optimized Bt cryIA(b) gene [bssyn]. This
figure shows that the full-length synthetic maize optimized
cryIA(b) gene sequence matches that of the native cryIA(b) gene
at about 2354 out of 3468 nucleotides (approximately 68%
homology).

Fig. 2 is a comparison of the truncated native Bt
cryIA(b) gene [BTHKURHD] and a truncated synthetic maize

optimized Bt gene [bssyn]. This figure shows that the
truncated synthetic maize optimized cryIA(b) gene sequence
matches that of the native cryIA(b) gene at about 1278 out of
1947 nucleotides (approximately 66% homology).

Fig. 3 is a comparison of the pure maize optimized Bt
gene sequence [syn1T.mze] with a truncated synthetic maize
optimized Bt gene [bssyn] and a full-length synthetic maize
optimized Bt gene modified to include restriction sites for
facilitating construction of the gene [synful.mod]. This
figure shows that the truncated synthetic maize optimized
cryIA(b) gene sequence matches that of the pure maize optimized
cryIA(b) gene at 1913 out of 1947 nucleotides (approximately
98% homology).

Fig. 4 is a comparison of a native truncated Bt
cryIA(b) gene [BTHKURHD] with a truncated synthetic cryIA(b)
gene described in Perlak et al., PNAS USA, 88:3324-3328 (1991)
[PMONBT] and a truncated synthetic maize optimized Bt gene
[bssyn]. This figure shows that the PMONBT gene sequence
matches that of the native cryIA(b) gene at about 1453 out of
1845 nucleotides (approximately 79% homology), while the
truncated synthetic maize optimized Bt cryIA(b) gene matches
the native cryIA(b) gene at about 1209 out of 1845 nucleotides
(approximately 66% homology).

Fig. 5 is a comparison of a truncated synthetic
cryIA(b) gene described in Perlak et al., PNAS USA,
88:3324-3328 (1991) [PMONBT] and a truncated synthetic maize
optimized Bt cryIA(b) gene [bssyn]. This figure shows that the

PMONBT gene sequence matches that of the truncated synthetic maize optimized Bt cryIA(b) gene at about 1410 out of 1845 nucleotides (approximately 77% homology).

Fig. 6 is a full-length, maize optimized CryIB gene.

Fig. 7 is a full-length, hybrid, partially maize optimized DNA sequence of a CryIA(b) gene which is contained in pCIB4434. The synthetic region is from nucleotides 1-1938 (amino acids 1-646), and the native region is from nucleotides 1939-3468 (amino acids 647-1155). The fusion point between the synthetic and native coding sequences is indicated by a slash (/) in the sequence.

Fig. 8 is a map of pCIB4434.

Fig. 9 is a full-length, hybrid, maize optimized DNA sequence encoding a heat stable CryIA(b) protein, contained in pCIB5511.

Fig. 10 is a map of pCIB5511.

Fig. 11 is a full-length, hybrid, maize optimized DNA sequence encoding a heat stable CryIA(b) protein, contained in pCIB5512.

Fig. 12 is a map of pCIB5512.

Fig. 13 is a full-length, maize optimized DNA sequence encoding a heat stable CryIA(b) protein, contained in pCIB5513.

Fig. 14 is a map of pCIB5513.

Fig. 15 is a full-length, maize optimized DNA sequence encoding a heat-stable CryIA(b) gene, contained in pCIB5514.

Fig. 16 is a map of pCIB5514.

Fig. 17 is a map of pCIB4418.

10

Fig. 18 is a map of pCIB4420.

Fig. 19 is a map of pCIB4429.

Fig. 20 is a map of pCIB4431.

Fig. 21 is a map of pCIB4428.

Fig. 22 is a map of pCIB4430.

Fig. 23A is a table containing data of cryIA(b) protein levels in transgenic maize.

Fig. 23B is a table which summarizes results of bioassays of Ostrinia and Diatraea on leaf material from maize progeny containing a maize optimized CryIA(b) gene.

Fig. 23C is a table containing data of cryIA(b) protein levels in transgenic maize.

Fig. 23D is a table which summarizes the results of bioassays of Ostrinia and Diatraea on leaf material from maize progeny containing a synthetic Bt. maize gene operably linked to a pith promoter.

Fig. 23E is a table containing data on expression of the cryIA(b) gene in transgenic maize using the pith-preferred promoter.

Fig. 24 is a complete genomic DNA sequence of a maize tryptophan synthase-alpha subunit gene. Introns, exons, transcription and translation starts, start and stop of cDNA are shown. $ = start and end of cDNA; +1 = transcription start; 73******* = primer extension primer; +1 = start of translation; +++ = stop codon; bp 1495-99 = CCAAT Box; bp 1593-1598 = TATAA Box; bp 3720-3725 = poly A addition site; ∤ above underlined sequences are PCR primers.

Figs. 25A, 25B, 25C and 25D are Northern blot analyses which show differential expression of the maize TrpA subunit gene in maize tissue at 2 hour, 4 hour, 18 hour, and 48 hour intervals, respectively, at -80°C with DuPont Cronex intensifying screens. P=pith; C=cob; BR=brace roots; ES=ear shank; LP=lower pith; MP=middle pith; UP=upper pith; S=seed; L=leaf; R=root; and P(upper left)=total pith.

Fig. 26 is a Northern blot analysis, the two left lanes of which show the maize TrpA gene expression in the leaf (L) and pith (P) of Funk inbred lines 211D and 5N984. The five right lanes indicate the absence of expression in Funk 211D seed total RNA. S(1, 2,3, 4 and 5)= seed at 1, 2, 3, 4 and 5 weeks post pollenation. L=leaf; P=pith; S#=seed # weeks post pollenation.

Fig. 27 is a Southern blot analysis of genomic DNA Funk line 211D, probed with maize TrpA cDNA 8-2 (pCIB5600), wherein B denotes BamHI, E denotes EcoRI, EV denotes EcoRV, H denotes HINDIII, and S denotes SacI. 1X, 5X and 10X denote reconstructed gene copy equivalents.

Fig. 28A is a primer extension analysis which shows the transcriptional start of the maize TrpA subunit gene and sequencing ladder. Lane +1 and +2 are 1X + 0.5X samples of primer extension reaction.

Fig. 28B is an analysis of RNase protection from +2 bp to +387 bp at annealing temperatures of 42°C, 48°C and 54°C, at a 16 hour exposure against film at -80°C with DuPont Cronex intensifying screens.

12

Fig. 29 is A map of the original Type II
pollen-specific cDNA clone. The subcloning of the three EcoRI
fragments into pBluescript vectors to create pCIB3168, pCIB3169
and II-.6 is illustrated.

Fig. 30 shows the DNA sequence of the maize
pollen-specific calcium dependent protein kinase gene cDNA, as
contained in the 1.0 kb and 0.5 kb fragments of the original
Type II cDNA clone. The EcoRI site that divides the 1.0 kb and
0.5 kb fragments is indicated. This cDNA is not full length,
as the mRNA start site maps 490 bp upstream of the end of the
cDNA clone.

Fig. 31 illustrates the tissue-specific expression of
the pollen CDPK mRNA. RNA from the indicated maize 211D
tissues was denatured, electrophoresed on an agarose gel,
transferred to nitrocellulose, and probed with the pollen CDPK
cDNA 0.5 kb fragment. The mRNA is detectable only in the
pollen, where a strong signal is seen.

Fig. 32 is an amino acid sequence comparison of the
pollen CDPK derived protein sequence and the rat protein kinase
2 protein sequence disclosed in Tobimatsu et al., J. Biol.
Chem. 263:16082-16086 (1988). The Align program of the DNAstar
software package was used to evaluate the sequences. The
homology to protein kinases occurs in the 5' two thirds of the
gene, i.e. in the 1.0 kb fragment.

Fig. 33 is an amino acid sequence comparison of the
pollen CDPK derived protein sequence and the human calmodulin
protein sequence disclosed in Fischer et al., J. Biol. Chem.

13

263:17055-17062 (1988). The homology to calmodulin occurs in the 3' one third of the gene, i.e. in the 0.5 kb fragment.

Fig. 34 is an amino acid sequence comparison of the pollen CDPK derived protein sequence and soybean CDPK. The homology occurs over the entire gene.

Fig. 35 illustrates the sequence of the maize pollen-specific CDPK gene. 1.4 kb of sequence prior to the mRNA start site is shown. The positions of the seven exons and six introns are depicted under the corresponding DNA sequence. The site of polyadenylation in the cDNA clone is indicated.

Fig. 36 is a map of pCIB4433.

Fig. 37 is a full-length, hybrid, maize-optimized DNA sequence encoding a heat stable cryIA(b) protein.

Fig. 38 is a map of pCIB5515.

<u>Description of the Sequences:</u>

Sequence 1 is the DNA sequence of a full-length native Bt cryIA(b) gene.

Sequence 2 is the DNA sequence of a full-length pure maize optimized synthetic Bt cryIA(b) gene.

Sequence 3 is the DNA sequence of an approximately 2 Kb truncated synthetic maize optimized Bt cryIA(b) gene.

Sequence 4 is the DNA sequence of a full-length synthetic maize optimized Bt cryIA(b) gene.

Sequence 5 is the DNA sequence of an approximately 2 Kb synthetic Bt gene according to Perlak et al.

<u>Detailed Description of the Invention</u>

The following definitions are provided in order to

14

provide clarity with respect to the terms as they are used in the specification and claims to describe the present invention.

Maize preferred codon: Preferred codon refers to the preference exhibited by a specific host cell in the usage of nucleotide codons to specify a given amino acid. The preferred codon for an amino acid for a particular host is the single codon which most frequently encodes that amino acid in that host. The maize preferred codon for a particular amino acid may be derived from known gene sequences from maize. For example, maize codon usage for 28 genes from maize plants are listed in Table 4 of Murray et al., Nucleic Acids Research, 17:477-498 (1989), the disclosure of which is incorporated herein by reference. For instance, the maize preferred codon for alanine is GCC, since, according to pooled sequences of 26 maize genes in Murray et al., supra, that codon encodes alanine 36% of the time, compared to GCG (24%), GCA (13%), and GCT (27%).

Pure maize optimized sequence: An optimized gene or DNA sequence refers to a gene in which the nucleotide sequence of a native gene has been modified in order to utilize preferred codons for maize. For example, a synthetic maize optimized Bt cryIA(b) gene is one wherein the nucleotide sequence of the native Bt cryIA(b) gene has been modified such that the codons used are the maize preferred codons, as described above. A pure maize optimized gene is one in which the nucleotide sequence comprises 100 percent of the maize preferred codon sequences for a particular polypeptide. For example, the pure

15

maize optimized Bt cryIA(b) gene is one in which the nucleotide sequence comprises 100 percent maize preferred codon sequences and encodes a polypeptide with the same amino acid sequence as that produced by the native Bt cryIA(b) gene. The pure nucleotide sequence of the optimized gene may be varied to permit manipulation of the gene, such as by altering a nucleotide to create or eliminate restriction sites. The pure nucleotide sequence of the optimized gene may also be varied to eliminate potentially deleterious processing sites, such as potential polyadenylation sites or intron recognition sites.

It is recognized that "partially maize optimized," sequences may also be utilized. By partially maize optimized, it is meant that the coding region of the gene is a chimeric (hybrid), being comprised of sequences derived from a native insecticidal gene and sequences which have been optimized for expression in maize. A partially optimized gene expresses the insecticidal protein at a level sufficient to control insect pests, and such expression is at a higher level than achieved using native sequences only. Partially maize optimized sequences include those which contain at least about 5% optimized sequences.

Full-length Bt Genes: Refers to DNA sequences comprising the full nucleotide sequence necessary to encode the polypeptide produced by a native Bt gene. For example, the native Bt cryIA(b) gene is approximately 3.5 Kb in length and encodes a polypeptide which is approximately 1150 amino acids in length. A full-length synthetic cryIA(b) Bt gene would be

16

at least approximately 3.5 Kb in length.

Truncated Bt Genes: Refers to DNA sequences comprising
less than the full nucleotide sequence necessary to encode the
polypeptide produced by a native Bt gene, but which encodes the
active toxin portion of the polypeptide. For example, a
truncated synthetic Bt gene of approximately 1.9 Kb encodes the
active toxin portion of the polypeptide such that the protein
product exhibits insecticidal activity.

Tissue-preferred promoter: The term "tissue-
preferred promoter" is used to indicate that a given regulatory
DNA sequence will promote a higher level of transcription of an
associated structural gene or DNA coding sequence, or of
expression of the product of the associated gene as indicated
by any conventional RNA or protein assay, or that a given DNA
sequence will demonstrate some differential effect; i.e., that
the transcription of the associated DNA sequences or the
expression of a gene product is greater in some tissue than in
all other tissues of the plant.

"Tissue-specific promoter" is used to indicate that a
given regulatory DNA sequence will promote transcription of an
associated coding DNA sequence essentially entirely in one or
more tissues of a plant, or in one type of tissue, e.g. green
tissue, while essentially no transcription of that associated
coding DNA sequence will occur in all other tissues or types of
tissues of the plant.

The present invention provides DNA sequences optimized
for expression in plants, especially in maize plants. In a

17

preferred embodiment of the present invention, the DNA sequences encode the production of an insecticidal toxin, preferably a polypeptide sharing substantially the amino acid sequence of an insecticidal crystal protein toxin normally produced by Bacillus thuringiensis. The synthetic gene may encode a truncated or full-length insecticidal protein. Especially preferred are synthetic DNA sequences which encode a polypeptide effective against insects of the order Lepidoptera and Coleoptera, and synthetic DNA sequences which encode a polypeptide having an amino acid sequence essentially the same as one of the crystal protein toxins of Bacillus thuringiensis variety kurstaki, HD-1.

The present invention provides synthetic DNA sequences effective to yield high expression of active insecticidal proteins in plants, preferably maize protoplasts, plant cells and plants. The synthetic DNA sequences of the present invention have been modified to resemble a maize gene in terms of codon usage and G+C content. As a result of these modifications, the synthetic DNA sequences of the present invention do not contain the potential processing sites which are present in the native gene. The resulting synthetic DNA sequences (synthetic Bt IP coding sequences) and plant transformation vectors containing this synthetic DNA sequence (synthetic Bt IP genes) result in surprisingly increased expression of the synthetic Bt IP gene, compared to the native Bt IP gene, in terms of insecticidal protein production in plants, particularly maize. The high level of expression

18

results in maize cells and plants that exhibit resistance to
lepidopteran insects, preferably European Corn Borer and
Diatrea saccharalis, the Sugarcane Borer.

The synthetic DNA sequences of the present invention
are designed to encode insecticidal proteins from Bacillus
thuringiensis, but are optimized for expression in maize in
terms of G+C content and codon usage. For example, the maize
codon usage table described in Murray et al., supra, is used to
reverse translate the amino acid sequence of the toxin produced
by the Bacillus thuringiensis subsp. kurstaki HD-1 cryIA(b)
gene, using only the most preferred maize codons. The reverse
translated DNA sequence is referred to as the pure maize
optimized sequence and is shown as Sequence 4. This sequence
is subsequently modified to eliminate unwanted restriction
endonuclease sites, and to create desired restriction
endonuclease sites. These modifications are designed to
facilitate cloning of the gene without appreciably altering the
codon usage or the maize optimized sequence. During the
cloning procedure, in order to facilitate cloning of the gene,
other modifications are made in a region that appears
especially susceptible to errors induced during cloning by the
polymerase chain reaction (PCR). The final sequence of the
maize optimized synthetic Bt IP gene is shown in Sequence 2. A
comparision of the maize optimized synthetic Bt IP gene with
the native kurstaki cryIA(b) Bt gene is shown in Fig. 1.

In a preferred embodiment of the present invention, the
protein produced by the synthetic DNA sequence is effective

19

against insects of the order <u>Lepidoptera</u> or <u>Coleoptera</u>. In a more preferred embodiment, the polypeptide encoded by the synthetic DNA sequence consists essentially of the full-length or a truncated amino acid sequence of an insecticidal protein normally produced by <u>Bacillus thuringiensis</u> var. <u>kurstaki HD-1</u>. In a particular embodiment, the synthetic DNA sequence encodes a polypeptide consisting essentially of a truncated amino acid sequence of the Bt CryIA(b) protein.

The insecticidal proteins of the invention are expressed in a plant in an amount sufficient to control insect pests, i.e. insect controlling amounts. It is recognized that the amount of expression of insecticidal protein in a plant necessary to control insects may vary depending upon species of plant, type of insect, environmental factors and the like. Generally, the insect population will be kept below the economic threshold which varies from plant to plant. For example, to control European corn borer in maize, the economic threshold is .5 eggmass/plant which translates to about 10 larvae/plant.

The methods of the invention are useful for controlling a wide variety of insects including but not limited to rootworms, cutworms, armyworms, particularly fall and beet armyworms, wireworms, aphids, corn borers, particularly European corn borers, sugarcane borer, lesser corn stalk borer, Southwestern corn borer, etc.

In a preferred embodiment of the present invention, the synthetic coding DNA sequence optimized for expression in

maize comprises a G+C percentage greater than that of the native cryIA(b) gene. It is preferred that the G+C percentage be at least about 50 percent, and more preferably at least about 60 percent. It is especially preferred that the G+C percent be about 64 percent.

In another preferred embodiment of the present invention, the synthetic coding DNA sequence optimized for expression in maize comprises a nucleotide sequence having at least about 90 percent homology with the "pure" maize optimized nucleotide sequence of the native _Bacillus thuringiensis_ cryIA(b) protein, more preferably at least about 95 percent homology, and most preferably at least about 98 percent.

Other preferred embodiments of the present invention include synthetic DNA sequences having essentially the DNA sequence of Sequence ID No. 4, as well as mutants or variants thereof; transformation vectors comprising essentially the DNA sequence of Sequence ID No. 4; and isolated DNA sequences derived from the plasmids pCIB4406, pCIB4407, pCIB4413, pCIB4414, pCIB4416, pCIB4417, pCIB4418, pCIB4419, pCIB4420, pCIB4421, pCIB4423, pCIB4434, pCIB4429, pCIB4431, pCIB4433. Most preferred are isolated DNA sequences derived from the plasmids pCIB4418 and pCIB4420, pCIB4434, pCIB4429, pCIB4431, and pCIB4433.

In order to construct one of the maize optimized DNA sequences of the present invention, synthetic DNA oligonucleotides are made with an average length of about 80 nucleotides. These oligonucleotides are designed to hybridize

21

to produce fragments comprising the various quarters of the truncated toxin gene. The oligonucleotides for a given quarter are hybridized and amplified using PCR. The quarters are then cloned and the cloned quarters are sequenced to find those containing the desired sequences. In one instance, the fourth quarter, the hybridized oligonucleotides are cloned directly without PCR amplification. Once all clones of four quarters are found which contain open reading frames, an intact gene encoding the active insecticidal protein is assembled. The assembled gene may then be tested for insecticidal activity against any insect of interest including the European Corn Borer (ECB) and the sugarcane borer. (Examples 5A and 5B, respectively). When a fully functional gene is obtained, it is again sequenced to confirm its primary structure. The fully functional gene is found to give 100% mortality when bioassayed against ECB. The fully functional gene is also modified for expression in maize.

The maize optimized gene is tested in a transient expression assay, e.g. a maize transient expression assay. The native Bt cryIA(b) coding sequence for the active insecticidal toxin is not expressed at a detectable level in a maize transient expression system. Thus, the level of expression of the synthesized gene can be determined. By the present methods, expression of a protein in a transformed plant can be increased at least about 100 fold to about 50,000 fold, more specifically at least about 1,000 fold to at least about 20,000 fold.

Increasing expression of an insecticial gene to an
effective level does not require manipulation of a native gene
along the entire sequence. Effective expression can be
achieved by manipulating only a portion of the sequences
necessary to obtain increased expression. A full-length, maize
optimized CryIA(b) gene may be prepared which contains a
protein of the native CryIA(b) sequence. For example, Figure 7
illustrates a full-length, maize optimized CryIA(b) gene which
is a synthetic-native hybrid. That is, about 2kb of the gene
(nucleotides 1-1938) is maize optimized, i.e. synthetic. The
remainder, C-terminal nucleotides 647-1155, are identical to
the corresponding sequence native of the CryIA(b) gene.
Construction of the illustrated gene is described in Example 6,
below.

It is recognized that by using the methods described
herein, a variety of synthetic/native hybrids may be
constructed and tested for expression. The important aspect of
hybrid construction is that the protein is produced in
sufficient amounts to control insect pests. In this manner,
critical regions of the gene may be identified and such regions
synthesized using preferred codons. The synthetic sequences
can be linked with native sequences as demonstrated in the
Examples below. Generally, N-terminal portions or processing
sites can be synthesized and substituted in the native coding
sequence for enhanced expression in plants.

In another embodiment of the present invention, the
maize optimized genes encoding cryIA(b) protein may be

23

manipulated to render the encoded protein more heat stable or
temperature stable compared to the native cryIA(b) protein. It
has been shown that the cryIA(b) gene found in _Bacillus_
thuringiensis kurstaki HD-1 contains a 26 amino acid deletion,
when compared with the cryIA(a) and cryIA(c) proteins, in the
-COOH half of the protein. This deletion leads to a
temperature-sensitive cryIA(b) protein. See M. Geiser, EP 0
440 581, entitled "Temperaturstabiles _Bacillus_
thuringiensis-Toxin". Repair of this deletion with the
corresponding region from the cryIA(a) or cryIA(c) protein
improves the temperature stability of the repaired protein.
Constructs of the full-length modified cryIA(b) synthetic gene
are designed to insert sequences coding for the missing amino
acids at the appropriate place in the sequence without altering
the reading frame and without changing the rest of the protein
sequence. The full-length synthetic version of the gene is
assembled by synthesizing a series of double-stranded DNA
cassettes, each approximately 300 bp in size, using standard
techniques of DNA synthesis and enzymatic reactions. The
repaired gene is said to encode a "heat stable" or
"temperature-stable" cryIA(b) protein, since it retains more
biological activity than its native counterpart when exposed to
high temperatures. Specific sequences of maize optimized, heat
stable cryIA(b) genes encoding temperature stable proteins are
set forth in Figs. 9, 11, 13, and 15, and are also described in
Example 7, below.

The present invention encompasses maize optimized

coding sequences encoding other polypeptides, including those

of other <u>Bacillus</u> <u>thuringiensis</u> insecticidal polypeptides or

insecticidal proteins from other sources. For example, cryIB

genes can be maize optimized, and then stably introduced into

plants, particularly maize. The sequence of a maize optimized

cryIB gene constructed in accordance with the present invention

is set forth in Fig. 6.

Optimizing a Bt IP gene for expression in maize using

the maize preferred codon usage according to the present

invention results in a significant increase in the expression

of the insecticidal gene. It is anticipated that other genes

can be synthesized using plant codon preferences to improve

their expression in maize or other plants. Use of maize codon

preference is a likely method of optimizing and maximizing

expression of foreign genes in maize. Such genes include genes

used as selectable or scoreable markers in maize

transformation, genes which confer herbicide resistance, genes

which confer disease resistance, and other genes which confer

insect resistance.

The synthetic cryIA(b) gene is also inserted into

<u>Agrobacterium</u> vectors which are useful for transformation of a

large variety of dicotyledenous plant species. (Example 44).

Plants stably transformed with the synthetic cryIA(b)

<u>Agrobacterium</u> vectors exhibit insecticidal activity.

The native Bt cryIA(b) gene is quite A+T rich. The G+C

content of the full-length native Bt cryIA(b) gene is

approximately 39%. The G+C content of a truncated native Bt

cryIA(b) gene of about 2 Kb in length is approximately 37%. In general, maize coding regions tend to be predominantly G+C rich. The modifications made to the Bt cryIA(b) gene result in a synthetic IP coding region which has greater than 50% G+C content, and has about 65% homology at the DNA level with the native cryIA(b) gene. The protein encoded by this synthetic CryIA(b) gene is 100% homologous with the native protein, and thus retains full function in terms of insect activity. The truncated synthetic CryIA(b) IP gene is about 2 Kb in length and the gene encodes the active toxin region of the native Bt kurstaki CryIA(b) insecticidal protein. The length of the protein encoded by the truncated synthetic CryIA(b) gene is 648 amino acids.

The synthetic genes of the present invention are useful for enhanced expression in transgenic plants, most preferably in transformed maize. The transgenic plants of the present invention may be used to express the insecticidal CryIA(b) protein at a high level, resulting in resistance to insect pests, preferably coleopteran or lepidopteran insects, and most preferably European Corn Borer (ECB) and Sugarcane Borer.

In the present invention, the DNA coding sequence of the synthetic maize optimized gene may be under the control of regulatory elements such as promoters which direct expression of the coding sequence. Such regulatory elements, for example, include monocot or maize and other monocot functional promoters to provide expression of the gene in various parts of the maize plant. The regulatory element may be constitutive. That is,

it may promote continuous and stable expression of the gene.
Such promoters include but are not limited to the CaMV 35S
promoter; the CaMV 19S promoter; A. tumefaciens promoters such
as octopine synthase promoters, mannopine synthase promoters,
nopaline synthase promoters, or other opine synthase promoters;
ubiquitin promoters, actin promoters, histone promoters and
tubulin promoters. The regulatory element may be a
tissue-preferential promoter, that is, it may promote higher
expression in some tissues of a plant than in others.
Preferably, the tissue-preferential promoter may direct higher
expression of the synthetic gene in leaves, stems, roots and/or
pollen than in seed. The regulatory element may also be
inducible, such as by heat stress, water stress, insect feeding
or chemical induction, or may be developmentally regulated.
Numerous promoters whose expression are known to vary in a
tissue specific manner are known in the art. One such example
is the maize phosphoenol pyruvate carboxylase (PEPC), which is
green tissue-specific. See, for example, Hudspeth, R.L. and
Grula, J.W., Plant Molecular Biology 12:579-589, 1989). Other
green tissue-specific promoters include chlorophyll a/b binding
protein promoters and RubisCO small subunit promoters.

The present invention also provides isolated and
purified pith-preferred promoters. Preferred pith-preferred
promoters are isolated from graminaceous monocots such as
sugarcane, rice, wheat, sorghum, barley, rye and maize; more
preferred are those isolated from maize plants.

In a preferred embodiment, the pith-preferred promoter

27

is isolated from a plant TrpA gene; in a most preferred embodiment, it is isolated from a maize TrpA gene. That is, the promoter in its native state is operatively associated with a maize tryptophan synthase-alpha subunit gene (hereinafter "TrpA"). The encoded protein has a molecular mass of about 38kD. Together with another alpha subnit and two beta subunits, TrpA forms a multimeric enzyme, tryptophan synthase. Each subunit can operate separately, but they function more efficiently together. TrpA catalyzes the conversion of indole glycerol phosphate to indole. Neither the maize TrpA gene nor the encoded protein had been isolated from any plant before Applicants' invention. The Arabidopsis thaliana tryptophan synthase beta subunit gene has been cloned as described Wright et al., The Plant Cell, 4:711-719 (1992). The instant maize TrpA gene has no homology to the beta subunit encoding gene.

The present invention also provides purified pollen-specific promoters obtainable from a plant calcium-dependent phosphate kinase (CDPK) gene. That is, in its native state, the promoter is operably linked to a plant CDPK gene. In a preferred embodiment, the promoter is isolated from a maize CDPK gene. By "pollen-specific," it is meant that the expression of an operatively associated structural gene of interest is substantially exclusively (i.e. essentially entirely) in the pollen of a plant, and is negligible in all other plant parts. By "CDPK," it is meant a plant protein kinase which has a high affinity for calcium, but not calmodulin, and requires calcium, but not calmodulin, for its

28

catalytic activity.

To obtain tissue-preferred or tissue specific promoters, genes encoding tissue specific messenger RNA (mRNA) can be obtained by differential screening of a cDNA library. For example, a pith-preferred cDNA can be obtained by subjecting a pith cDNA library to differential screening using cDNA probes obtained from pith and seed mRNA. See, Molecular Cloning, A Laboratory Manual, Sambrook et al. eds. Cold Spring Harbor Press: New York (1989).

Alternately, tissue specific promoters may be obtained by obtaining tissue specific proteins, sequencing the N-terminus, synthesizing oligonucleotide probes and using the probes to screen a cDNA library. Such procedures are exemplified in the Experimental section for the isolation of a pollen specific promoter.

The scope of the present invention in regard to the pith-preferred and pollen-specific promoters encompasses functionally active fragments of a full-length promoter that also are able to direct pith-preferred or pollen-specific transcription, respectively, of associated structural genes. Functionally active fragments of a promoter DNA sequence may be derived from a promoter DNA sequence, by several art-recognized procedures, such as, for example, by cleaving the promoter DNA sequence using restriction enzymes, synthesizing in accordance with the sequence of the promoter DNA sequence, or may be obtained through the use of PCR technology. See, e.g. Mullis et al., Meth. Enzymol. 155:335-350 (1987); Erlich (ed.), PCR

29

Technology, Stockton Press (New York 1989).

Further included within the scope of the instant invention are pith-preferred and pollen-specific promoters "equivalent" to the full-length promoters. That is, different nucleotides, or groups of nucleotides may be modified, added or deleted in a manner that does not abolish promoter activity in accordance with known procedures.

A pith-preferred promoter obtained from a maize TrpA gene is shown in Fig. 24. Those skilled in the art, with this sequence information in hand, will recognize that pith-preferred promoters included within the scope of the present invention can be obtained from other plants by probing pith libraries from these plants with probes derived from the maize TrpA structural gene. Probes designed from sequences that are highly conserved among TrpA subunit genes of various species, as discussed generally in Example 17, are preferred. Other pollen-specific promoters, which in their native state are linked to plant CDPK genes other than maize, can be isolated in similar fashion using probes derived from the conserved regions of the maize CDPK gene to probe pollen libraries.

In another embodiment of the present invention, the pith-preferred or pollen-specific promoter is operably linked to a DNA sequence, i.e. structural gene, encoding a protein of interest, to form a recombinant DNA molecule or chimeric gene. The phrase "operably linked to" has an art-recognized meaning; it may be used interchangeably with "operatively associated

30

with,""linked to," or "fused to".

The structural gene may be homologous or heterologous
with respect to origin of the promoter and/or a target plant
into which it is transformed. Regardless of relative origin,
the associated DNA sequence will be expressed in the
transformed plant in accordance with the expression properties
of the promoter to which it is linked. Thus, the choice of
associated DNA sequence should flow from a desire to have the
sequence expressed in this fashion. Examples of heterologous
DNA sequences include those which encode insecticidal proteins,
e.g. proteins or polypeptides toxic or inhibitory to insects or
other plant parasitic arthropods, or plant pathogens such as
fungi, bacteria and nematodes. These heterologous DNA
sequences encode proteins such as magainins, Zasloff, PNAS USA,
84:5449-5453 (1987); cecropins, Hultmark et al., Eur. J.
Biochem. 127:207-217 (1982); attacins, Hultmark et al., EMBO J.
2:571-576 (1983); melittin, gramicidin S, Katsu et al.,
Biochem. Biophys. Acta, 939:57-63 (1988); sodium channel
proteins and synthetic fragments, Oiki et al. PNAS USA,
85:2395-2397 (1988); the alpha toxin of Staphylococcus aureusm
Tobkes et al., Biochem., 24:1915-1920 (1985); apolipoproteins
and fragments thereof, Knott et al., Science 230:37 (1985);
Nakagawa et al., J. Am. Chem. Soc., 107:7087 (1985);
alamethicin and a variety of synthetic amphipathic peptides,
Kaiser et al., Ann. Rev. Biophys. Biophys. Chem. 16:561-581
(1987); lectins, Lis et al., Ann. Rev. Biochem., 55:35-68
(1986); protease and amylase inhibitors; and insecticidal

31

proteins from <u>Bacillus</u> <u>thuringiensis</u>, particularly the delta-endotoxins from <u>B.</u> <u>thuringiensis</u>; and from other bacteria or fungi.

In a preferred embodiment of the invention, a pith-preferred promoter obtained from a maize TrpA subunit gene or pollen-specific promoter obtained from a maize CDPK gene is operably linked to a heterologous DNA sequence encoding a <u>Bacillus</u> <u>thuringiensis</u> ("B.t.") insecticidal protein. These proteins and the corresponding structural genes are well known in the art. <u>See</u>, Hofte and Whiteley, <u>Microbiol. Reviews</u>, 53:242-255 (1989).

While it is recognized that any promoter capable of directing expression can be utilized, it may be preferable to use heterologous promoters rather than the native promoter of the protein of interest. In this manner, chimeric nucleotide sequences can be constructed which can be determined based on the plant to be transformed as well as the insect pest. For example, to control insect pests in maize, a monocot or maize promoter can be operably linked to a Bt protein. The maize promoter can be selected from tissue-preferred and tissue-specific promoters such as pith-preferred and pollen-specific promoters, respectively as disclosed herein.

In some instances, it may be preferred to transform the plant cell with more than one chimeric gene construct. Thus, for example, a single plant could be transformed with a pith-preferred promoter operably linked to a Bt protein as well as a pollen-specific promoter operably linked to a Bt protein.

The transformed plants would express Bt proteins in the plant pith and pollen and to a lesser extent the roots, outer sheath and brace roots.

For various other reasons, particularly management of potential insect resistance developing to plant expressed insecticidal proteins, it is beneficial to express more than one insecticidal protein (IP) in the same plant. One could express two different genes (such as two different <u>Bacillus thuringiensis</u> derived delta-endotoxins which bind different receptors in the target insect's midgut) in the same tissues, or one can selectively express the two toxins in different tissues of the same plant using tissue specific promoters. Expressing two Bt genes (or any two insecticidal genes) in the same plant using three different tissue specific promoters presents a problem for production of a plant expressing the desired phenotype. Three different promoters driving two different genes yields six different insecticidal genes that need to be introduced into the plant at the same time. Also needed for the transformation is a selectable marker to aid in identification of transformed plants. This means introducing seven different genes into the plant at the same time. It is most desired that all genes, especially the insecticidal genes, integrate into the plant genome at the same locus so they will behave as a single gene trait and not as a multiple gene trait that will be harder to track during breeding of commercial hybrids. The total number of genes can be reduced by using differential tissue specific expression of the different

33

insecticidal proteins.

For example, by fusing cryIA(b) with the pollen and PEP carboxylase promoters, one would obtain expression of this gene in green tissues and pollen. Fusing a pith-preferred promoter with the cryIB delta endotoxin from <u>Bacillus</u> <u>thuringiensis</u> would produce expression of this insecticidal protein most abundantly in the pith of a transformed plant, but not in seed tissues. Transformation of a plant with three genes, PEP carboxylase/cryIA(b), pollen/cryIA(b), and pith/cryIB produces a plant expressing two different Bt insecticidal endotoxins in different tissues of the same plant. CryIA(b) would be expressed in the "outside" tissues of a plant (particularly maize), that is, in those tissues which European corn borer feeds on first after hatching. Should ECB prove resistant to cryIA(b) and be able to burrow into the stalk of the plant after feeding on leaf tissue and/or pollen, it would then encounter the cryIB delta-endotoxin and be exposed to a second insecticidal component. In this manner, one can differentially express two different insecticdal components in the same plant and decrease the total number of genes necessary to introduce as a single genetic unit while at the same time providing protection against development of resistance to a single insecticidal component.

Likewise, a plant may be transformed with constructs encoding more than one type of insecticidal protein to control various insects. Thus, a number of variations may be constructed by one of skill in the art.

34

The recombinant DNA molecules of the invention may be prepared by manipulating the various elements to place them in proper orientation. Thus, adapters or linkers may be employed to join the DNA fragments. Other manipulations may be performed to provide for convenient restriction sites, removal of restriction sites or superfluous DNA. These manipulations can be performed by art-recognized methods. See, Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, second edition, 1989. For example, methods such as restriction, chewing back or filling in overhangs to provide blunt ends, ligation of linkers, complementary ends of the DNA fragments can be provided for joining and ligation. See, Sambrook et al., supra.

Other functional DNA sequences may be included in the recombinant DNA molecule, depending upon the way in which the molecule is to be incorporated into the target plant genome. For instance, in the case of Agrobacterium-mediated transformation, if Ti- or the Ri- plasmid is used to transform the plant cells, the right and left borders of the T-DNA of the Ti- and Ri- plasmid will be joined as flanking regions to the expression cassette. Agrobacterium tumefaciens-mediated transformation of plants has been described in Horsch et al., Science, 225:1229 (1985); Marton, Cell Culture Somatic Cell Genetics of Plants, 1:514-521 (1984); Hoekema, In: The Binary Plant Vector System Offset-Drukkerij Kanters B.V., Alblasserdam, 1985, Chapter V Fraley, et al., Crit. Rev. Plant Sci., 4:1-46; and An et al., EMBO J., 4:277-284 (1985).

35

The recombinant DNA molecules of the invention also can include a marker gene to facilitate selection in recombinant plant cells. Examples of markers include resistance to a biocide such as an antibiotic, e.g. kanamycin, hygromycin, chloramphenicol, paramomycin, methotrexate and bleomycin, or a herbicide such as imidazolones, sulfonylureas, glyphosate, phosphinothricin, or bialaphos. Marker genes are well known in the art.

In another embodiment of the present invention, plants stably transformed with a recombinant DNA molecule or chimeric gene as described hereinabove are provided. The resultant transgenic plant contains the transformed gene stably incorporated into its genome, and will express the structural gene operably associated to the promoter in the respective fashion.

Transgenic plants encompassed by the instant invention include both monocots and dicots. Representative examples include maize, tobacco, tomato, cotton, rape seed, soybean, wheat, rice, alfalfa, potato and sunflower. [others?]. Preferred plants include maize, particularly inbred maize plants.

All transformed plants encompassed by the instant invention may be prepared by several methods known in the art. A. tumefaciens-mediated transformation has been disclosed above. Other methods include direct gene transfer into protoplasts, Paszkowski et al., EMBO J., 12:2717 (1984); Loerz et al., Mol. Gen. & Genet., 1199:178 (1985); Fromm et al.,

<u>Nature</u> 319:719 (1986); microprojectile bombardment, Klein et

al., <u>Bio/Technology</u>, 6:559-563 (1988); injection into

protoplasts, cultured cells and tissues, Reich et al.,

<u>Bio/Technology</u>, 4:1001-1004 (1986); or injection into

meristematic tissues or seedlings and plants as described by De

La Pena et al., <u>Nature</u>, 325:274-276 (1987); Graves et al.,

<u>Plant Mol. Biol.</u>, 7:43-50 (1986); Hooykaas-Van Slogteren et

al., <u>Nature</u>, 311:763-764 (1984); Grimsley et al.,

<u>Bio/Technology</u>, 6:185 (1988); and Grimsley et al., <u>Nature</u>,

325:177 (1988); and electroporation, WO92/09696.

The expression pattern of a structural gene operatively

associated with an instant tissue-preferred or tissue-specific

promoter in a transformed plant containing the same is critical

in the case where the structural gene encodes an insecticidal

protein. For example, the instantly disclosed pith-preferred

expression pattern will allow the transgenic plant to tolerate

and withstand pathogens and herbivores that attack primarily

the pith, but also the brace roots, outer sheath and leaves of

the plant since the protein will be expressed to a lesser

extent but still in an insect controlling amount in these plant

parts, but yet in the case of both types of promoters, will

leave the seed of the plant unaffected.

<u>EXAMPLES</u>

The following examples further describe the materials

and methods used in carrying out the invention. They are

offered by way of illustration, and not by way of limitation.

EXAMPLE 1: General Methods

DNA manipulations were done using procedures that are standard in the art. These procedures can often be modified and/or substituted without substantively changing the result. Except where other references are identified, most of these procedures are described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, second edition, 1989.

Synthesis of DNA oligomers:

DNA oligomers which are from about twenty to about ninety, preferably from about sixty to about eighty nucleotides in length, are synthesized using an Applied Biosystems model 380B DNA synthesizer and standard procedures. The oligomers are made using the updated SSCAF3 cycle on a 0.2 µmole, wide pore, small scale ABI column. The end procedure is run trityl off and the oligomer is cleaved from the column using the 380B's automatic cleavage cycle. The oligomers are then deblocked in excess ammonium hydroxide (NH_4OH) at 55°C for 8-12 hours. The oligomers are then dried in an evaporator using nitrogen gas. After completion, the oligomers are resuspended in 0.25 - 0.5 ml of deionized water.

Purification of synthetic oligomers:

An aliquot of each oligomer is mixed with an equal volume of blue dye\formamide mix with the final solution containing 0.05% bromophenol blue, 0.05% xylene cyanol FF, and

25 % formamide. This mixture is heated at 95°C for 10 minutes

to denature the oligomers. Samples are then applied to a 12 %

polyacrylamide-urea gel containing 7 M urea (Sambrook et al.).

After electrophoresis at 300-400 volts for 3-4 hours using a

Vertical Slab Gel Unit (Hoefer Scientific Instruments, San

Francisco, CA), UV shadowing is used to locate the correct

sized fragment in the gel which was then excised using a razor

blade. The purified gel fragment is minced and incubated in

0.4 M LiCl, 1 mM EDTA (pH 8) buffer overnight at 37°C.

    Either of two methods is used to separate the oligomers

from the polyacrylamide gel remnants: Gene\X 25 µM porous

polyethylene filter units or Millipore's ultrafree-MC 0.45 µM

filter units. The purified oligomers are ethanol precipitated,

recovered by centrifuging in a microfuge for 20 min at 4°C, and

finally resuspended in TE (10 mM Tris, 1 mM EDTA, pH 8.0).

Concentrations are adjusted to 50 ng\µl based on absorption

readings at 260 nm.

Kinasing oligomers for size determinations:

    To check the size of some of the oligomers on a

sequencing gel, kinase labeling reactions are carried out using

purified synthetic oligomers of each representative size:

40mers, 60mers, 70mers, 80mers, and 90mers. In each 20 µl

kinasing reaction, one pmole of purified oligomer is used in a

buffer of 7.0 mM Tris pH 7.5, 10 mM KCl, 1 mM MgCl2), 0.5 mM

DTT, 50 µg/ml BSA, 3000 µCi (3 pmoles) of 32P-gammaATP, and

8 units of T4 polynucleotide kinase. The kinase reaction is

incubated for 1 hour at 37°C, followed by a phenol\chloroform

extraction and three ethanol precipitations with glycogen as

carrier (Tracy, Prep. Biochem. 11:251-268 (1981).

Two gel loadings (one containing 1000 cpm, the other

containing 2000 cpm) of each reaction are prepared with 25%

formamide, 0.05% bromophenol blue, and 0.05% xylene cyanol FF.

The kinased oligomers are boiled for 5 minutes before loading

on a 6 % polyacrylamide, 7 M urea sequencing gel (BRL Gel

Mix TM6, BRL, Gaithersburg, MD). A sequencing reaction of

plasmid pUC18 is run on the same gel to provide size markers.

After electrophoresis, the gel is dried and exposed to

diagnostic X-ray film (Kodak, X-OMAT AR). The resulting

autoradiograph shows all purified oligomers tested to be of the

correct size. Oligomers which had not been sized directly on

the sequencing gel are run on a 6 % polyacrylamide, 7 M urea

gel (BRL Gel Mix TM6), using the sized oligomers as size

markers. All oligomers are denatured first with 25 % formamide

at 100°C for 5 minutes before loading on the gel. Ethidium

bromide staining of the polyacrylamide gel allows all the

oligomers to be visualized for size determination.

Hybridizing oligomers for direct cloning:

Oligomers to be hybridized are pooled together (from 1

µg to 20 µg total DNA) and kinased at 37°C for 1 hour in 1X

Promega ligation buffer containing 30 mM Tris-HCl pH 7.8, 10 mM

MgCl2, 10 mM DTT, and 1 mM dATP. One to 20 units of T4

polynucleotide kinase is used in the reaction, depending on the

amount of total DNA present. The kinasing reactions are
stopped by placing the reaction in a boiling water bath for
five minutes. Oligomers to form the 5' termini of the
hybridized molecules are not kinased but are added to the
kinased oligomers along with additional hybridization buffer
after heating. The pooled oligomers are in a volume of 50-100
ul with added hybridization buffer used to adjust the final
salt conditions to 100 mM NaCl, 120 mM Tris pH 7.5, and 10 mM
MgCl2. The kinased and non-kinased oligomers are pooled
together and heated in a boiling water bath for five minutes
and allowed to slowly cool to room temperature over a period of
about four hours. The hybridized oligomers are then
phenol\chloroform extracted, ethanol precipitated, and
resuspended in 17 µl of TE (10 mM Tris, 1 mM EDTA, pH 8.0).
Using this 17 µl, a ligation reaction with a final volume of 20
µl is assembled (final conditions = 30 mM Tris-HCl pH 7.8, 10
mM MgCl2, 10 mM DTT, 1 mM ATP, and 3 units of T4 DNA ligase
(Promega, Madison WI). The ligation is allowed to incubate for
about 2 hours at room temperature. The hybridized\ligated
fragments are generally purified on 2% Nusieve gels before
and\or after cutting with restriction enzymes prior to cloning
into vectors. A 20 µl volume ligation reaction is assembled
using 100 ng to 500 ng of each fragment with approximate
equimolar amounts of DNA in 30 mM Tris-HCl pH 7.8, 10 mM MgCl2,
10 mM DTT, 1 mM ATP, and 3 units of T4 DNA ligase (Promega,
Madison, WI). Ligations are incubated at room temperature for
2 hours. After ligation, DNA is transformed into frozen

competent E. coli cells using standard procedures (Sambrook et
al.) and transformants are selected on LB-agar (Sambrook et
al.) containing 100 µg/ml ampicillin (see below).

PCR Reactions for Screening clones in E. coli:

    E. coli colonies which contain the correct DNA insert
are identified using PCR (see generally, Sandhu et al.,
BioTechniques 7:689-690 (1989)). Using a toothpick, colonies
are scraped from an overnight plate and added to a 20 µl to 45
µl PCR reaction mix containing about 50 pmoles of each
hybridizing primer (see example using primers MK23A28 and
MK25A28 to select orientation of SacII fragment in pHYB2#6),
200 µm to 400 mM of each dNTP, and 1X reaction buffer (Perkin
Elmer Cetus, Norwalk, CT). After boiling the E. coli\PCR mix
in a boiling water bath for 10 minutes, 5 µl of Taq polymerase
(0.5 units)(Perkin Elmer Cetus, Norwalk, Conn.) in 1X reaction
buffer is added. The PCR reaction parameters are generally set
with a denaturing step of 94°C for 30 seconds, annealing at
55°C for 45 seconds, and extension at 72°C for 45 seconds for
30 to 36 cycles. PCR reaction products are run on agarose or
Nusieve agarose (FMC) gels to detect the correct fragment size
amplified.

Ligations:

    Restriction enzyme digested fragments are either
purified in 1% LGT (low gelling temperature agarose, FMC), 2%
Nusieve (FMC), or 0.75% agarose using techniques standard in

42

the art. DNA bands are visualized with ethidium bromide and
bands are recovered from gels by excision with a razor blade.
Fragments isolated from LGT are ligated directly in the LGT.
Ten microliters of each recovered DNA fragment is used to
assemble the ligation reactions, producing final ligation
reaction volumes of about 23 µl. After excision with a razor
blade, the recovered gel bands containing the desired DNA
fragments are melted and brought to 1X ligase buffer and 3
units of T4 DNA ligase (Promega) are added as described above.
Fragments isolated from either regular agarose or Nusieve
agarose are purified from the agarose using ultrafree-MC 0.45
µM filter units (Millipore) and the fragments are ligated as
described above. Ligation reactions are incubated at room
temperature for two hours before transforming into frozen
competent E. coli cells using standard procedures (Sambrook et
al.).

Transformations:

        Frozen competent E. coli cells of the strain DH5alpha
or HB101 are prepared and transformed using standard procedures
(Sambrook et al.). E. Coli "SURE" competent cells are obtained
from Stratagene (La Jolla, CA). For ligations carried out in
LGT agarose, after ligation reactions are complete, 50 mM CaCl2
is added to a final volume of about 150 µl and the solution
heated at approximately 65°C for about 10 minutes to completely
melt the agarose. The solution is then mixed and chilled on
ice for about 10 minutes before the addition of about 200 µl of

competent cells which had been thawed on ice. This mixture is allowed to incubate for 30 minutes on ice. The mixture is next heat shocked at 42°C for 60 seconds before chilling on ice for two minutes. Next, 800 μl of SOC media (20% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, adjusted to pH 8 with 5 N NaOH, 20 mM MgCl2:MgSO4 mix, and 20 mM glucose; Sambrook et al.) is added and the cells are incubated at 37°C with shaking for about one hour before plating on selective media plates. Plates typically are L-agar (Sambrook et al.) containing 100 μg/ml ampicillin.

When ligations are carried out in a solution without agarose, typically 200 μl of frozen competent E. coli cells (strain DH5alpha (BRL, Gaithersburg, MD or Sure cells, Stratagene, La Jolla, CA) are thawed on ice and 5 μl of the ligation mixture added. The reaction is incubated on ice for about 45 to 60 minutes, the cells are then heat shocked at 42° for about 90 seconds. After recovery at room temperature for about 10 minutes, 800 μl of SOC medium is added and the cells are then incubated 1 hour at 37°C with shaking and plated as above.

When screening for inserts into the beta-galactosidase gene in some of the standard vectors used, 200 μl of the recovered transformation mixture is plated on LB-agar plates containing 0.008% X-gal, 80 μM IPTG, and 100 μg/ml ampicillin (Sambrook et al.). The plates are incubated at 37° overnight to allow selection and growth of transformants.

44

Miniscreening DNA:

Transformants from the selective media plates are grown and their plasmid structure is examined and confirmed using standard plasmid mini-screen procedures (Sambrook et al.). Typically, the "boiling" procedure is used to produce small amounts of plasmid DNA for analysis (Sambrook et al.). Alternatively, an ammonium acetate procedure is used in some cases. This procedure is a modification of that reported by Shing-yi Lee et al., Biotechniques 9:676-679 (1990).

1) Inoculate a single bacterial colony from the overnight selection plates into 5 ml (can be scaled down to 1 ml) of TB (Sambrook et al.) medium and grow in the presence of the appropriate antibiotic.

2) Incubate on a roller at 37°C overnight.

3) Collect 5 ml of bacterial cells in a plastic Oakridge tube and spin for 5 min. at 5000 rpm in a Sorvall SS-34 rotor at 4°C.

4) Remove the supernatant.

5) Resuspend the pellet in 1 ml of lysis buffer (50 mM glucose, 25 mM Tris-HCl[pH 8.0], 10 mM EDTA and 5 mg/ml lysozyme), vortex for 5 seconds, and incubate at room temperature for 5 min.

6) Add 2 ml of freshly prepared alkaline solution (0.2 N NaOH, 1% sodium dodecyl sulfate), tightly secure lid, mix by inverting 5 times and place tube in an ice-water bath for 5 min.

7) Add 1.5 ml of ice-cold 7.5 M ammonium acetate (pH

7.6) to the solution, mix by inverting the tube gently 5 times and place on an ice-water bath for 5 min.

8) Centrifuge mixture at 9000 rpm for 10 min. at room temperature.

9) Transfer clear supernatant to a 15 ml Corex tube and add 0.6 volumes of isopropanol (approx. 2.5 ml). Let sit at room temperature for 10 min.

10) Centrifuge the mixture at 9000 rpm for 10 min. at room temperature and discard the supernatant.

11) Resuspend the pellet in 300 ul of TE buffer. Add 6 ul of a stock of RNase A & T1 (made as a 200 ul solution by adding 180 ul of RNase A [3254 Units/mg protein, 5.6 mg protein/ml] and 20 ul of RNase T1[481 Units/ug protein, 1.2 mg protein/ml]). These stocks may be purchased from USB (US Biochemical). Transfer to a microcentrifuge tube and incubate at 37°C for 15 min.

12) Add 75 ul of distilled water and 100 ul of 7.5 M ammonium acetate and incubate in an ice-water bath for 10 min.

13) Centrifuge the mixture at 14,000 rpm for 10 min. in a Beckman microfuge at 4°C.

14) Precipitate by adding 2.5 volumes of 100% EtOH (approx. 1 ml) and incubate in an ice-water bath for 10 min.

15) Spin at 14,000 rpm for 10 min. in a microfuge.

16) Wash pellet with 70% ethanol (using 0.5 ml-1 ml). Dry the pellet and resuspend in 100 µl of 1X New England Biolabs restriction enzyme Buffer 4 [20 mM Tris-HCl(pH 7.9), 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT].

Measure concentration and check purity by spectrophotometry at
absorbances 260 and 280 nm.

For a more rapid determination as to whether or not a
particular bacterial colony harbored a recombinant plasmid, a
PCR miniscreen procedure is carried out using a modification of
the method described by (Sandhu, G.S. et al., 1989,
BioTechniques, 7:689-690). Briefly, the following mixture is
prepared:

100 µl primer mix above, 20 µM each primer,

100 µl dNTP mix (2.5 mM each)

100 µl 10X AmpliTaq buffer (Perkin-Elmer Cetus, 1X
buffer = 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl2, and
0.01% gelatin)

700 µl deionized water.

20 µl of the above mixture is put into a a 0.5 ml
polyproplyene PCR tube. A transformed bacterial colony is
picked with a toothpick and resuspended in the mixture. The
tube is put in a boiling water bath for 10 minutes and then
cooled to room temperature before adding 5 µl of the mix
described below:

265 µl deionized water

30 µl 10X Amplitaq buffer (Perkin-Elmer Cetus, 1X
buffer = 10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM MgCl2, and
0.01% gelatin)

7.5 µl Taq polymerase

The samples are overlaid with 50 µl of mineral oil and PCR is carried out for 30 cycles using the following parameters:

denature: 94° for 1 min

anneal: 55° for 1 min

extend: 72° for 45 seconds.

After PCR amplification, 1 µl of loading dye (30% glycerol, 0.25% Bromophenol blue, 0.25% xylene cyanol) is added to the whole reaction and 20 µl of the mixture is loaded on a 2% Nusieve, 1% agarose gel to see if there is a PCR product of the expected size.

This procedure is used as an initial screen. Minipreps are subsequently carried out to confirm the structure of the plasmid and its insert prior to sequencing.

EXAMPLE 2: AMPLIFICATION AND ASSEMBLY OF EACH QUARTER
Cloning fragments of the synthetic Bt cryIA(b) gene:

The synthetic gene was designed to be cloned in four pieces, each roughly one quarter of the gene. The oligomers for each quarter were pooled to either be assembled by PCR, hybridization, or a combination of hybridization followed by PCR amplification as described elsewhere. Synthetic quarters were pieced together with overlapping restriction sites Aat II, NcoI, and Apa I between the 1st and 2nd, 2nd and 3rd, and 3rd and 4th quarters respectively.

Each quarter of the gene (representing about 500 bp)

48

was assembled by hybridizing the appropriate oligomers and
amplifying the desired fragment using PCR primers specific for
the ends of that quarter. Two different sets of PCR reactions
employing two sets of slightly different primers were used.
The PCR products of the two reactions were designed to be
identical except that in the first reaction there was an
additional AATT sequence at the 5' end of the coding region and
in the second reaction there was an AGCT sequence at the 3' end
of a given quarter. When the products of the two reactions for
a particular quarter were mixed (after removing the polymerase,
primers and incomplete products), denatured, and subsequently
re-annealed, a certain ratio (theoretically 50%) of the
annealed product should have non-homologous overhanging ends.
These ends were designed to correspond to the "sticky ends"
formed during restriction digestion with EcoRI at the 5' end
and Hind III at the 3' end of the molecule. The resulting
molecules were phosphorylated, ligated into an EcoRI/HindIII
digested and phosphatased Bluescript vector, and transformed
into frozen competent E. coli strain DH5alpha. After
selection, the E. coli colonies containing the desired fragment
are identified by restriction digest patterns of the DNA.
Inserts representing parts of the synthetic gene are
subsequently purified and sequenced using standard procedures.
In all cases, clones from multiple PCR reactions are generated
and sequenced. The quarters are then joined together using the
unique restriction sites at the junctions to obtain the
complete gene.

49

Cloned quarters are identified by mini-screen
procedures and the gene fragment sequenced. It is found that
errors are frequently introduced into the sequence, most
probably during the PCR amplification steps. To correct such
errors in clones that contain only a few such errors,
hybridized oligomers are used. Hybridized fragments are
digested at restriction enzyme recognition sites within the
fragment and cloned to replace the mutated region in the
synthetic gene. Hybridized fragments range from 90 bp in
length (e.g. the region that replaces the fragment between the
Sac II sites in the 2nd quarter) to the about 350 bp 4th
quarter fragment that replaces two PCR induced mutations in the
4th quarter of the gene.

Due to the high error rate of PCR, a plasmid is
designed and constructed which allows the selection of a cloned
gene fragment that contains an open reading frame. This
plasmid is designed in such a manner that if an open reading
frame is introduced into the cloning sites, the transformed
bacteria could grow in the presence of kanamycin. The
construction of this vector is described in detail below. This
selection system greatly expedites the progress by allowing one
to rapidly identify clones with open reading frames without
having to sequence a large number of independent clones. The
synthetic quarters are assembled in various plasmids, including
BSSK (Stratagene; La Jolla, Ca), pUC18 (Sambrook et al.), and
the Km-expression vector. Other suitable plasmids, including
pUC based plasmids, are known in the art and may also be used.

Complete sequencing of cloned fragments, western blot analysis
of cloned gene products, and insect bioassays using European
corn borer as the test insect verify that fully functional
synthetic Bt cryIA(b) genes have been obtained.

Construction of the Km-expression vector to select open reading
frames:

The Km-expression vector is designed to select for fragments of
the synthetic gene which contain open-reading frames. PCR
oligomers are designed which allow the fusion of the NPTII gene
from Tn5 starting at nucleotide 13 (Reiss et al., EMBO J.
3:3317-3322 (1984)) with pUC18 and introduce useful restriction
sites between the DNA segments. The polylinker region contains
restriction sites to allow cloning various synthetic Bt IP
fragments in-frame with the Km gene. The 88 bp 5' oligomer
containing the polylinker region is purified on a 6%
polyacrylamide gel as described above for the oligomer PAGE
purification. A PCR reaction is assembled with a 1 Kb Bgl
II\Sma I template fragment which contains the NPT II gene
derived from Tn5. The PCR reaction mix contains 100 ng of
template with 100 pmols of oligomers KE72A28 and KE74A28 (see
sequences below), 200 nM dNTP, and 2.5 Units of Taq polymerase
all in a 50 µl volume with an equal volume of mineral oil
overlaid. Sequences of the primers are:

KE74A28

5'-GCAGATCTGG ATCCATGCAC GCCGTGAAGG GCCCTTCTAG AAGGCCTATC

GATAAAGAGC TCCCCGGGGA TGGATTGCAC GCAGGTTC-3'

KE72A28

5'-GCGTTAACAT GTCGACTCAG AAGAACTCGT CAAGAAGGCG-3'

The PCR parameters used are: 94°C for 45 seconds (sec), 55°C for 45 sec, and 72°C for 55 sec with the extension at step 3 for 3 sec for 20 cycles. All PCR reactions are carried out in a Perkin-Elmer Cetus thermocycler. The amplified PCR product is 800 bp and contains the polylinker region with a translational start site followed by unique restriction sites fused in-frame with the Km gene from base #13 running through the translational terminator. pUC:KM74 is the Km-expression cassette that was assembled from the 800 bp Bgl II\Sal I polylinker/Km fragment cloned in the PUC18 vector. The lacZ promoter allows the Km gene to be expressed in E. coli. pUC:KM74 derivatives has to first be plated on LB-agar plates containing 100 µg/ml ampicillin to select transformants which can subsequently be screened on LB-agar plates containing 25 µg/ml kanamycin/IPTG. Synthetic Bt IP gene fragments are assembled from each quarter in the Km-cassette to verify cloning of open-reading-frame containing fragments pieces. The first ECB active synthetic Bt IP gene fragment, pBt:Km#6, is a Bt IP gene that shows Km resistance. This fragment is subsequently discovered to contain mutations in the 3rd and 4th quarter which are later repaired.

EXAMPLE 2A: SYNTHESIS AND CLONING OF THE FIRST QUARTER OF THE
             SYNTHETIC GENE [base pairs 1 to 550]

The following procedures are followed in order to clone

the first quarter of the synthetic DNA sequence encoding a

synthetic Bt cryIA(b) gene. The same procedures are

essentially followed for synthesis and cloning of the other

quarters, except as noted for primers and restriction sites.

Template for Quarter 1: Mixture of equal amounts of purified

oligomers U1-U7 and L1 to L7

PCR Primers:

Forward:

P1 (a):   5'-GTCGACAAGG ATCCAACAAT GG-3'

P1 (b):   5'-AATTGTCGAC AAGGATCCAA CAATGG-3'

Reverse:

P2 (a):   5'-ACACGCTGAC GTCGCGCAGC ACG-3'

P2 (b):   5'-AGCTACACGC TGACGTCGCG CAG-3'

Primer pair A1: P1(b) + P2(a)

Primer pair A2: P1(a) + P2(b)

The PCR reaction containing the oligomers comprising

the first quarter of the synthetic maize-optimized Bt IP gene

is set up as follows:

53

200 ng oligo mix (all oligos for the quarter mixed in equal amounts based on weight)

10 μl of primer mix (1:1 mix of each at 20 μM; primers are described above)

5 μl of 10X PCR buffer

PCR buffer used may be either

(a) 1X concentration = 10 mM KCl, 10 mM (NH4)2SO4, 20 mM Tris-HCl, pH 8.0, 2 mM MgSO4, and 0.1% Triton X-100), or

(b) 1X concentration = 10mM Tris-HCl pH 8.3, 50 mM KCl 1.5mM $MgCl_2$, 0.01% wt/vol gelatin.

Components are mixed, heated in a boiling water bath for 5 minutes, and incubated at 65°C for 10 minutes.

Next, the following reagents are added:

8 μl of dNTPs mixture (final concentration in the reaction = 0.2 mM each)

5 units polymerase.

The final reaction volume is 50 microliters.

Oligomers are then incubated for 3 min at 72°C and then a PCR cycle is run. The PCR reaction is run in a Perkin Elmer thermocycler on a step cycle protocol as follows:

denaturation cycle : 94° for 1 minute

annealing cycle : 60° for 1 minute

extension cycle : 72° for 45 seconds (+ 3 sec per cycle)

number of cycles: 15

54

After the reaction is complete, 10 µl of the PCR
reaction is loaded on a 2% Nusieve-GTG (FMC), 1% agarose
analytical gel to monitor the reaction. The remaining 40 µl is
used to clone the gene fragments as described below.

PCR Products

The termini of the double stranded PCR product
corresponding to the various primer pairs are shown (only upper
strand):

A1 AATTGTCGAC_____GCGTGT   (554 bp) first qtr.

A2   GTCGAC_____GCGTGTAGCT (554 bp) first qtr.

Hybridization

40 µl of each of the PCR reactions described above is
purified using a chromaspin 400 column (Clonetech, Palo Alto,
CA) according to manufacturers directions. Five µg of carrier
DNA was added to the reactions before loading on the column.
(This is done for most of the cloning. However, in some
reactions the PCR reactions are phenol:chloroform extracted
using standard procedures (Sambrook et al.) to remove the Taq
polymerase and the PCR generated DNA is recovered from the
aqueous phase using a standard ethanol precipitation
procedure.) The carrier DNA does not elute with the PCR
generated fragments. The A1 and A2 reaction counterparts for
each quarter are mixed, heated in a boiling water bath for 10

minutes and then incubated at 65°C overnight. The reactions
are then removed from the 65° bath and ethanol precipitated
with 1 µl (20 µg) of nuclease free glycogen (Tracy, Prep.
Biochem. 11:251-268 (1981) as carrier. The pellet is
resuspended in 40 µl of deionized water.

Phosphorylation reaction

The phosphorylation reaction is carried out as follows:

40 µl DNA

2.5 µl 20 mM ATP

0.5 µl 10X BSA/DTT (1X = 5 mM DTT, 0.5 mg/ml BSA)

1.0 µl 10X polynucleotide kinase buffer (1X = 70
mM Tris.HCl,

pH 7.6, 0.1 M KCl, 10 mM MgCl2)

2.0 µl polynucleotide kinase (New England Biolabs,
20 units).

Incubation is for 2 hours at 37°C.

The reaction is then extracted one time with a 1:1
phenol:chloroform mixture, then once with chloroform and the
aqueous phase ethanol precipitated using standard procedures.
The pellet is resuspended in 10 µl of TE.

Restriction Digests

20 µg of Bluescript vector (BSSK+, Stratagene, La
Jolla, CA)

10 µl 10 X restriction buffer (1X = 20 mM Tris-HCl pH

56

8.0, 10 mM MgCl2, 100 mM NaCl)

     5 μl Eco RI (New England Biolabs) 100 units

     5 μl Hind III (New England Biolabs) 100 units

     Final reaction volume is 100 μl.

     Incubation is for 3 hours at 37°.

When completed, the reaction is extracted with an equal volume
of phenol saturated with TE (10 mM Tris.HCl pH 8.0 and 10 mM
EDTA). After centrifugation, the aqueous phase was extracted
with an equal volume of 1:1 mixture of (TE saturated)
phenol:chloroform (the "chloroform" is mixed in a ratio of 24:1
chloroform:isoamyl alcohol), and finally the aqueous phase from
this extraction is extracted with an equal volume of
chloroform. The final aqueous phase is ethanol precipitated
(by adding 10 μl of 3 M sodium acetate and 250 μl of absolute
ethanol, left at 4° for 10 min and centrifuged in a microfuge
at maximum speed for 10 minutes. The pellet is rinsed in 70%
ethanol and dried at room temperature for 5-10 minutes and
resuspended in 100 μl of 10 mM Tris.HCl (pH 8.3).

Phosphatase reaction

     Vector DNA is routinely treated with phosphatase to
reduce the number of colonies obtained without an insert. Calf
intestinal alkaline phosphatase is typically used (Sambrook et
al.), but other phosphatase enzymes can also be used for this
step.

     Typical phosphatase reactions are set up as below:

     90 μl of digested DNA described above

10 µl of 10X Calf intestinal alkaline phosphatase

buffer (1X=50 mM Tris-HCl (pH 8.3), 10 mM MgCl2, 1 mM ZnCl2, 10

mM spermidine)

1 µl (1 unit) of calf intestinal alkaline phosphatase

(CIP, Boehringer Mannheim, Indianapolis, IN)

Incubation is at 37°C for 1 hour.

The DNA is then gel purified (on a 1% low gelling

temperature (LGT) agarose gel) and the pellet resuspended in 50

µl TE. After electrophoresis, the appropriate band is excised

from the gel using a razor blade, melted at 65° for 5 minutes

and diluted 1:1 with TE. This solution is extracted twice with

phenol, once with the above phenol:chloroform mixture, and once

with chloroform. The final aqueous phase is ethanol

precipitated and resuspended in TE buffer.

Ligation:

To ligate fragments of the synthetic gene into vectors,

the following conditions are typically used.

5 µl of phosphorylated insert DNA

2 µl of phosphatased Eco RI/Hind III digested

Bluescript vector heated at 65° for 5 minutes, then cooled

1 µl 10X ligase buffer (1X buffer=30 mM Tris.HCl (pH

7.8), 10 mM MgCl2, 10 mM DTT, 1 mM ATP)

1 µl BSA (1 mg/ml)

1 µl ligase (3 units, Promega, Madison, Wisc.)

Ligase reactions are typically incubated at 16°

overnight or at room temperature for two hours.

58

Transformation:

Transformation of ligated DNA fragments into E. coli is performed using standard procedures (Sambrook et al.) as described above.

Identification of recombinants

White or light blue colonies resulting from overnight incubation of transformation plates are selected. Plasmids in the transformants are characterized using standard mini-screen procedures (Sambrook et al.) or as described above. One of the three procedures listed below are typically employed:

(1) boiling DNA miniprep method

(2) PCR miniscreen

(3) Ammonium acetate miniprep.

The restriction digest of recombinant plasmids believed to contain the first quarter is set up as follows:

(a) Bam HI/Aat II digest: 10 µl DNA + 10 µl 1X New England Biolabs restriction enzyme Buffer 4

0.5 µl Bam HI (10 units)

0.5 µl Aat II (5 units)

Incubation is for about 2 hours at 37°C.

Clones identified as having the desired restriction pattern are next digested with Pvu II and with Bgl II in separate reactions. Only clones with the desired restriction patterns with all three enzyme digestions are carried further for sequencing.

Sequencing of cloned gene fragments:

59

Sequencing is performed using a modification of Sanger's dideoxy chain termination method (Sambrook et al.) using double stranded DNA with the Sequenase 2 kit (United States Biochemical Corp., Cleveland, OH). In all, six first quarter clones are sequenced. Of the clones sequenced, only two clones designated pQA1 and pQA5 are found to contain only one deletion each. These deletions are of one base pair each located at position 452 in pQA1 and position 297 in pQA5.

Plasmid pQA1 is used with pP1-8 (as described below) to obtain a first quarter with the expected sequence.

EXAMPLE 2B: SYNTHESIS AND CLONING OF THE SECOND QUARTER [base pairs 531 to 1050]

Template: oligomers U8-U14 and L8-L14

PCR Primers:

forward:

P3 (a):   5'-GCTGCGCGAC GTCAGCGTGT TCGG-3'

P3 (b):   5'-AATTGCTGCG CGACGTCAGC GTG-3'

Reverse:

P4 (a):   5'-GGCGTTGCCC ATGGTGCCGT ACAGG-3'

P4 (b):   5'-AGCTGGCGT TGCCCATGGT GCCG-3'

Primer pair B1: P3(b) + P4(a)

Primer pair B2: P3(a) + P4(b)

PCR Products

B1 AATTGCTGCG_____AACGCC   (524 bp) second

quarter

B2   GCTGCG_____AACGCCAGCT (524 bp)

        Hybridization, PCR amplification, spin column size

fractionation, and cloning of this gene fragment in Bluescript

digested with Eco RI/Hind III are performed as described above

for the first quarter (Example 2A). The PCR product for this

quarter is about 529 bp in size representing the second quarter

of the gene (nucleotides 531 to 1050). Transformation is into

frozen competent E. coli cells (DH5alpha) using standard

procedures described above (Sambrook et al.)

Miniscreen of pQB clones:

        Miniprep DNA is prepared as described above and

digested with (a) Aat II/Nco I, (b) Pvu II and (c) with Bgl I

to confirm the structure insert in the vector before

sequencing.

        Sequencing is performed as described above using the

dideoxy method of Sanger (Sambrook et al.).

        A total of thirteen clones for this quarter are

sequenced. The second quarter consistently contains one or

more deletions between position 884 and 887. In most cases the

G at position 884 is deleted.

        Plasmid pQB5 had only one deletion at position 884.

This region lies between two Sac II sites (positions 859 and

949). Correction of this deletion is described in Example 3.

Clones of the first half (1-1050 bp).

A fragment for cloning the first half (quarters 1 and 2) of the synthetic Bt maize gene as a single DNA fragment is obtained by restriction digestion of the product of a PCR reaction comprising the first quarter and the second quarter. Restriction endonuclease Aat II is used to cut the DNA (following phenol extraction and ethanol precipitation) in a 20 µl reaction. 15 µl of each of the Aat II digested quarters is mixed and ligated (in a 50 µl volume by adding 5 µl of 10X ligase buffer, (1X=30 mM Tris-HCl pH 7.8, 10 mM MgCl2, 10 mM DTT, 1 mM ATP) 14 µl of deionized water and 1 µl of T4 DNA ligase, 3 units, Promega, Madison, WI) at room temperature for 2 hr. The result is an about 1 kb fragment as judged by electrophoresis on a 1% agarose gel run using standard conditions (Sambrook et al.) Ten µl of the ligation product is amplified by PCR using conditions described previously except that only 5 cycles were run.

Primer Pair: HA= P1(a) + P4(b)

Primer Pair: HB= P1(b) + P4(a)

The product of these reactions is cloned into Bluescript (Stratagene, La Jolla, CA) as described for the individual quarters. This procedure is only done once i.e., all insert DNA is obtained in a particular region from a single PCR reaction.

Thirty-six colonies are miniscreened with Sal I digests

and Pvu II digests. All except 4 contain an insert of

approximately 1 kb in size of which at least 20 contain the

correct Pvu II digestion pattern. Eight of these clones are

selected for sequence analysis. One of the clones, P1-8, has

the desired sequence between the Eco NI site (396 bp) and the

Dra III site (640 bp). This clone is used to obtain a plasmid

with the desired sequence up to the Dra III site (640 bp) in

the second quarter with pQA1 (first quarter with a deletion at

position 452 bp described previously.)

EXAMPLE 2C: CLONING AND SYNTHESIS OF THIRD
            QUARTER [base pairs 1021 to 1500]

Template: Oligos U15-U20 and L15-L21

PCR primers:

forward

P5 (a):   5'-TTCCCCCTGT ACGGCACCAT GGGCAACGCC GC-3'

P5 (b):   5'-AATTGTACGG CACCATGGGC AAC-3'

reverse

P6 (a):   5'-GAAGCCGGGG CCCTTCACCA CGCTGG-3'

P6 (b):   5'-AGCTGAAGCC GGGGCCCTTC ACC-3'

Primer pair C1: P5(b) + P6(a)

Primer pair C2: P5(a) + P6(b)

PCR Product:

C1   AATTGTACGG_____GGCTTC   (475 bp) 3d qtr

C2 TTCCCCTGTACGG_____GGCTTCAGCT (484 bp) 3d qtr

PCR reactions, spin column recovery of the correct sized DNA fragment, and ligation into vectors are performed as described above (Example 2A) using a Bluescript vector cut with Eco RI and Hind III. The approximately 479 base pair PCR product represents the third quarter of the synthetic gene (NT 1021 - 1500).

Transformation into frozen competent E. coli strain DH5alpha cells, selection and identification of transformants, characterization of transformants by mini-screen procedures, and sequencing of the synthetic gene fragment in the vector are all as described above.

Mini screen of pQC clones:

The third quarter is miniscreened using standard procedures (Sambrook et al.). Miniprep DNA is cut with (a) Nco I/Apa I and (b) with Pvu II. Clones containing the correct restriction digest patterns are sequenced using standard procedures. A total of 22 clones of the third quarter are sequenced. Three major deletion "hotspots" in the third quarter are identified (a) at position 1083 (b) between position 1290-1397 and (c) between positions 1356-1362. In all clones except one, pQC8, there is also consistently an insertion of a C at position 1365. In addition to these mutations, the third quarter clones contain a large number of other apparently random deletions. The common factor to the

three mutational "hotspots" in the third quarter and the one in

the second quarter is that these regions are all flanked on

either side by sequences that are about 80% C+G. Other regions

containing 5 to 9 C-Gs in a row are not affected. The

oligomers in U15, U16, U18, U19, L15, L16, L18 and L19 are

redesigned to reduce the C+G content in these regions. Five

clones each from PCR reaction using the modified oligomers are

sequenced.

Plasmid pQCN103 has the correct sequence for the third

quarter except for a change at position 1326. This change,

which substitutes a G for a C, results in the substitution of

one amino acid (leucine) for the original (phenylalanine).

EXAMPLE 2D: SYNTHESIS AND CLONING OF FOURTH
            QUARTER [base pairs 1480 to 1960]

The fourth quarter of the gene is obtained from a clone

which is originally designed to comprise the third and fourth

quarters of the gene. The "second half" of the synthetic gene

is obtained from PCR reactions to fuse the third and fourth

quarters. These reactions are run with PCR primers P5(a) and

P6(a) described above for the third quarter and primers P7(a)

and P8(a) (described below). The reverse primer is modified to

include a Sac I site and a termination codon. Separate

reactions for each quarter are run for 30 cycles using the

conditions described above. The two quarters are joined

together by overlapping PCR and subsequently digested with

restriction enzymes Nco I and Sac I. The resulting 953 bp

fragment is cloned directionally into pCIB3054, which has been

65

cut with Nco I/Sac I and treated with alkaline phosphatase.

pCIB3054 is constructed by inserting intron #9 of PEPcarboxylase (PEPC ivs #9) in the unique Hpa I site of pCIB246 (described in detail in Example 4) pCIB246 is cut with HpaI and phosphatased with CIP using standard procedures described in Example 2A. PEPC ivs #9 is obtained by PCR using pPEP-10 as the template. pPEP-10 is a genomic subclone containing the entire maize PEP carboxylase gene encoding the C_4 photosynthetic enzyme, plus about 2.2 Kb of 5'-flanking and 1.8 Kb of 3'-flanking DNA. The 10Kb DNA is ligated in the HindIII site of pUC18. (Hudspeth et al., <u>Plant Molecular Biology</u>, 12: 576-589 (1989). The forward PCR primer used to obtain the PEPCivs#9 is GTACAAAAACCAGCAACTC and the reverse primer is CTGCACAAAGTGGAGTAGT. The PCR product is a 108 bp fragment containing only the PEPcarboxylase intron #9 sequences. The PCR reaction is extracted with phenol and chloroform, ethanol precipitated phosphorylated with polynucleotide kinase and treated with T4 polymerase to fill in the 3' nontemplated base addition found in PCR products (Clark, J.M., <u>Nucleic Acid Research</u>, 16: 9677-9686 (1988)) using standard procedures. The kinased fragment is blunt-end cloned into the HpaI site of pCIB246, using standard procedures described earlier.

Amplification and Assembly of the Fourth Quarter

<u>Template:</u> U21-U26 and L22-L28

<u>PCR primers</u>

FORWARD

P7 (a):   5'-TGGTGAAGGG CCCCGGCTTC ACCGG-3'

REVERSE

P8 (a):   5'-ATCATCGATG AGCTCCTACA CCTGATCGAT GTGGTA-3'

PRIMER PAIR 4: P7(a) + P8(a)

PRIMER PAIR 3: P5(A) + P6(a)

Primer pair for overlapping PCR : P7(a) + P8(a)

PCR Product

fourth quarter: GGTGAA_____ATCAGGAGCTCATCGATGAT

(484 bp) third quarter: TTCCCCCTGTA--------------TTCACCGG

(484 bp) second half: GGTGAA-------CATGATGAT (953 bp)

Four positive clones are identified by plasmid
miniscreen and are subsequently sequenced using standard
procedures.

Plasmid Bt.P2 #1 contains approximately the correct
fourth quarter sequence except for two mutations. These
mutations are at position 1523 (substituting an A for a G,
resulting in an amino acid change which substitutes a His for
an Arg) and at position 1634 (substituting a T for a C,
resulting in an amino acid substitution of a Ser for a Thr).

Plasmid Bt.P2 #1 is used in the construction of
pCIB4414 described below. (The mistakes are ultimately
corrected by hybridizing all the oligos of the fourth quarter,
digesting with Apa I/Bst E II and replacing that region in
pCIB4414. Therefore, only sequences from position 1842-1960
remain of Bt.P2#1 in the final construct.)

EXAMPLE 3: ASSEMBLY AND REPAIR OF THE FINAL SYNTHETIC GENE

The synthetic maize optimized Bt cryIA(b) gene is designed to be cloned in quarters. Using the PCR technique, however, results in mutations, which in most cases are deletions resulting in frameshift mutations. Plasmids containing individual quarters are therefore sequenced and the correct parts ligated together using standard procedures.

After obtaining first and second quarter clones with almost the desired sequence, plasmids pEB1Q#4 and pEB1Q#5 are constructed to obtain the desired sequence of the synthetic Bt gene up to the Dra III site at the base pair position 634 (this mutation destroys the Dra III site). The pEB1Q constructs are made by ligating a 3.9 Kb Eco NI\Bam HI fragment from pP1-8 with a 400 bp fragment from pQA1. pEB1Q#5 has the desired sequence up to the Dra III site, but pEB1Q#4 has a mutation at base pair position 378.

Plasmids p1H1M4 and p1H1M5 are constructed to repair the Dra III site in pEB1Q#4 and pEB1Q#5. Plasmids p1H1M#4 and #5 are made by ligating a 3.5 Kb Nco I\Aat II fragment from pEB1Q#4 and #5 respectively, with a 500 bp Nco I\Aat II fragment from pQB5. Plasmid p1H1M5 contains a mutation between the Sac II sites at position 884 in the second quarter of the synthetic Bt gene. Plasmid p1H1M4 contains the additional mutation as described in its precursor construct pEB1Q#4.

The Sac II site in the Bluescript vector region of p1H1M4 is deleted by cutting p1H1M4 with Not I and Sac I and

converting these sites to blunt ends using T4 DNA polymerase under standard conditions before ligating this 3.9 Kb fragment to make p1H1M4^S. Deleting the Sac II site in the vector region allows the 90 bp Sac II fragment with the mutation at position 884 in the 2nd quarter of p1H1M4^S to be removed prior to replacement with a 90 bp Sac II fragment. Oligomers U\L 12 and 13 are kinased and hybridized (described above) before cutting with Sac II and isolating a 90 bp fragment on a 2% Nusieve gel. The Sac II fragment is ligated into the about 3.8 Kb Sac II cut p1H1M4^S vector which has been phosphatased with CIP. The repaired Sac II construct is called pHYB2#6. The orientation of the Sac II fragment in pHYB2#6 is detected by PCR screening as described earlier using the following primers:

MK23A28 = 5'-GGGGCTGCGGATGCTGCCCT-3'

MK25A28 = 5'-GAGCTGACCCTGACCGTGCT-3'

MK26A28 = 5'-CACCTGATGGACATCCTGAA-3'

Running the PCR reactions with 50 pmoles of primers MK23A28 and MK25A28 produces an approximate 180 bp fragment, indicating the inserted fragment bounded by the Sac II sites in pHYB2#6 is in the correct orientation. Using primers MK25A28 and MK26A28 in the PCR screening acts as the negative control producing an approximate 180 bp fragment only in constructs containing the Sac II bounded fragment in the wrong orientation. pHYB2#6 sequence is determined using standard procedures.

69

pHYB2#6 has one mutation at position 378 which needed to be repaired to obtain a first quarter containing the desired sequence.

Plasmid p1HG#6 contains the desired sequence for the entire first half of the synthetic Bt gene. p1HG#6 is made from a 3.4 Kb Aat II\Nco I fragment of p1H1M5#2 ligated to a 500 bp Aat II\Nco I fragment from pHYB2#6.

To identify clones or partial clones of the synthetic gene which contain open reading frames, the kanamycin selection vector (described above) is used. The fourth quarter of the synthetic Bt gene is the first put into the kanamycin cassette. pKM74-4 contains the approximately 500 bp Apa I\Cla I fragment from plasmid BtP2 (which had been previously transformed into a dam- E. coli strain (PO-100) to be able to cut with Cla I), ligated to pUC:KM74 cut with Apa I\Cla I. Plasmid pKM74-4 displays kanamycin resistance but is later found to contain two substitution mutations at positions 1523 and 1634 (mutations are described above in the section on cloning the fourth quarter; they are substitutions, not deletions or insertions).

The correct first half of the synthetic Bt gene from plasmid p1HG#6 is inserted into plasmid pKM74-4. The resulting plasmid, called pKm124, is made from the about 3.9 Kb Apa I\Bam HI fragment derived from pKM74-4 ligated to 1 Kb Apa I\Bam HI fragment from p1HG#6. pKm124 shows kanamycin resistance. This plasmid contains the first, second, and fourth quarters of the synthetic gene forming a single open reading frame.

The third quarter of the synthetic gene is next cloned

70

into pKM124. The first functional clone, in plasmid pBt:Km#6,

is a functional copy of the truncated synthetic cryIA(b) gene

in the Km-cassette which displays kanamycin resistance but

which contains deletion mutations between the third and fourth

quarters. Plasmid pBt:Km#6 is made from the approximately 5 Kb

Apa I\Nco I pKM124 vector fragment ligated to the approximately

500 bp Apa I\Nco I fragment from pQCN103 (pQCN103 contains a

mismatch mutation at position 1326 which is repaired later).

Contaminating nuclease activity appears to have deleted the Apa

I site between the third and fourth quarters in pBt:Km#6. The

Bt gene encoded by the synthetic gene in plasmid pBT:Km#6 has

about 50-60 % of the native proteins' activity against ECB.

The 2 Kb Sma I\Bam HI fragment from pBt:Km#6 is inserted into a

35S:expression cassette to make a plasmid called 35S:Bt6.

Two functional synthetic Bt clones, each with

mutations, are initially obtained: plasmids pBT:KM#6 and

pCIB4414. pCIB4414, which is 100% active in insect bioassays

against European corn borer compared with the native gene,

contains substitution mutations in the third and fourth

quarters at positions 1323, 1523, and 1634.

pCIB4414 is constructed from two plasmids, MG3.G4#18

and 1HG which is described above. MG3.G4#18 is obtained by

cloning the Apa I/Kpn I fragment in plasmid Bt.P2#1 into

pQCN103 (using those same restriction sites). This produces a

plasmid containing the third and fourth quarters of the gene.

The first half of the synthetic gene from plasmid 1HG is cut

with Bam HI and Nco I and moved into MG3.G4#18 (containing the

third and fourth quarters of the gene). The resulting plasmid, pCIB4414, contains a functional version of the synthetic gene. While being functional, the synthetic gene in this plasmid contains three errors; position 1326 (G substituted for a C), position 1523 (substitute A for a G), and at position 1634 (substitution of a T for a C).

The fourth quarter in pCIB4414 is replaced with a 354 bp fourth quarter Apa I\Bst E II fragment obtained from hybridizing, ligating, and restriction cleaving fourth quarter oligomers as described earlier, and isolating the fragment from a 2% Nusieve agarose gel. pCIB4408 is a synthetic Bt gene clone obtained by replacing the fourth quarter fragment in pCIB4414 with the hybridized fourth quarter fragment. To insert the CaMV 35S promoter in front of the synthetic Bt gene, pCIB4406 is made from a 4 Kb Eco NI\Kpn I fragment from plasmid p35SBt6 and 1.8 Kb Eco NI\Kpn I fragment from pCIB4408.

pCIB4406 is 100% active (as compared with the protein from the native gene) against ECB but contains the substitution mutation in the third quarter of the synthetic gene at position 1323 resulting in an amino acid substitution of a leucine for a phenylalanine. Plasmid pBS123#13 is used to repair this mutation.

The third quarter fragment in plasmid pBS123#13 is made from an approximately 479 bp hybridized oligomer generated fragment. Third quarter oligomers U15-U20 and L15-L21 are kinased, hybridized, and ligated as described above. PCR reactions are carried out as described above with primers P5(a)

72

and P6(b) for 15 cycles. The PCR product is treated with proteinase K at a final concentration of about 50 µg\ml in an approximate 95 µl volume for 30 minutes at 37°C followed by 10 minutes at 65°C (Crowe et al., Nucleic Acid Research 19:184, 1991.) Subsequently, the product is phenol\chloroform extracted and ethanol precipitated using standard procedures before cutting with restriction enzymes Apa I and Nco I.

The approximate 450 bp Apa I\Nco I PCR fragment is ligated to the 3.8 Kb Apa I\Nco I vector fragment from p1HG#6 to make pBS123#13. Plasmid pBS123#13 contains the desired sequence for the third quarter of the maize optimized cryIA(b) gene from position 1319 at the Nsp I site through the Apa I site at position 1493. This 170 bp Nsp I\Apa I fragment from pBS123#13 is used in the fully active synthetic cryIA(b) gene in plasmid pCIB4418.

Western Blot Analysis:

Western blot analyses of various transformants are performed using crude extracts obtained from E. coli grown on selective plates. Using a toothpick, cultures are scraped from fresh plates containing the transformants of interest which have been grown overnight at 37°C. The positive control for expression of the Bt gene in E. coli was a construct called pCIB3069 which contains the native Bt-k gene fused with the plant expressible CaMV 35S promoter. pCIB3069 also contains the 35S promoter operably linked to the hygromycin resistance gene, 35S promoter, with Adh intron #1 operably linked to the

GUS gene, and 35S promoter operably linked to a gene coding for the production of the native Bt cryIA(b) IP. A negative control of E. coli which does not contain a Bt gene is also included in the analyses. Cultures are resuspended in 100 µl of loading buffer containing 62 mM Tris-HCl pH 6.8, 1% SDS, 0.0025% bromophenol blue, 10% glycerol and 7.5% mercaptoethanol. After heating the mixtures at 95°C for 10 minutes, the preparations are sonicated for 1-3 seconds. The debris is centrifuged in a microfuge at room temperature for about 5 minutes and 10 to 15 µl of each sample is loaded onto an acrylamide gel with a 10% running gel below a 6% stacking gel (Laemmli, Nature 227;680-685(1970)). After electrophoresis overnight at 10 mAmps, proteins are transferred from the gel to an Immobilon membrane (Millipore). The transfer is done using an electrophoretic Blotting Unit (American BioNuclear, Emeryville, CA) in transfer buffer (20 mM Tris, 150 mM glycine, and 20% methanol) for 1.5 hours at 450 mAmps.

Buffers for western blotting included:

            Blocking buffer: 2% Tween-20

                               30 mM Tris-HCl pH 10.2

                               150 mM NaCl

      Wash buffer: 0.05% Tween-20

                        30 mM Tris-HCl pH 10.2

                        150 mM NaCl

      Developing buffer: 100 mM Tris-HCl pH 9.6

100 mM NaCl

10 mM MgCl2

After transfer is complete, the membrane is incubated
for about ten minutes in the blocking buffer. Three 15 minute
washes with wash buffer are done before the first antibody
treatment. The first antibody is an immunoaffinity purified
rabbit or goat antibody prepared using the CryIA(b) protein as
the antigen (Ciba-Geigy, RTP, N.C.; Rockland Inc.,
Gilbertsville, PA.; and Berkeley Antibody CO., Richmond, CA.).
The cryIA(b) specific antibody is treated immediately before
use with E. coli lysate from Bio-Rad in a 1 ml volume with 5 µg
of antibody, 50 µl of E. coli lysate in the wash buffer
solution. This mixture is incubated for 1 hour at room
temperature before diluting it 1 to 30 for a final dilution of
1:6000 with wash buffer. Incubation of the membrane with the
first antibody is at room temperature for 1.5 hours.

Three 10 minute washes are done between the 1st and 2nd
antibody treatments. The second antibody is either rabbit
anti-goat or goat anti-rabbit/alkaline phosphatase conjugate
(Sigma, St. Louis, MO.). Incubation with the alkaline
phosphatase conjugate is carried out at room temperature for
one hour using a 1 to 6000 dilution in wash buffer. Six 10
minute washes are done between the second antibody
treatment and developing the western blot. The western blot is
developed in 100 ml of developing buffer with 440 µl of
nitroblue tetrazolium in 70% dimethyl formamide (75 mg\ml), and

330 μl of 5-bromo-4-chloro-indolyl-phosphate in 100% dimethyl
formamide (50 mg\ml). After developing for 15 to 30 minutes,
the membrane is washed in water and air dried.

EXAMPLE 4: CONSTRUCTION OF TRANSFORMATION VECTORS

Construction of pCIB710 and derivatives.

    CaMV 35S Promoter Cassette Plasmids pCIB709 and pCIB710
are constructed as shown in Rothstein et al., Gene 53:153-161
(1987). pCIB710 contains CaMV promoter and transcription
termination sequences for the 35S RNA transcript [Covey et al.,
Nucl. Acids. Res., 9:6735-6747 (1981)]. A 1149 bp BglII
restriction fragment of CaMV DNA [bp 6494-7643 in Hohn et al.,
Current Topics in Microbiology and Immunology, 96:194-220 and
Appendices A to G (1982)] is isolated from CaMV DNA by
preparative agarose gel electrophoresis as described earlier
The fragment is mixed with BamHI-cleaved plasmid pUC19 DNA,
treated with T4 DNA ligase, and transformed into E. coli.
(Note the BamHI restriction site in the resulting plasmid is
destroyed by ligation of the BglII cohesive ends to the BamHI
cohesive ends.)

    The resulting plasmid, called pUC19/35S, is then used
in oligonucleotide-directed in-vitro mutagenesis to insert the
BamHI recognition sequence GGATCC immediately following CaMV
nucleotide 7483 in the Hohn reference. The resulting plasmid,
pCIB710, contains the CaMV 35S promoter region and
transcription termination region separated by a BamHI
restriction site. DNA sequences inserted into this BamHI site

will be expressed in plants by these CaMV transcription regulation sequences. (Also note that pCIB710 does not contain any ATG translation initiation codons between the start of transcription and the BamHI site).

pCIB710 is modified to produce pCIB709 by inserting a Bam HI fragment containing the coding sequence for hygromycin phosphotransferase from pLG90 [Rothstein et al., Gene, 53:153-161 (1987)] in the Bam HI site.

pCIB709 is modified to produce pCIB996 by removing the ATG just upstream from the initiation codon of the hygromycin phosphotranserase gene using standard mutagenesis techniques while inserting a Bgl II restriction site at this location. The resulting plasmid, pCIB996, is further modified to remove the Bam HI, Sma I and Bgl II sites in the 5' untranslated leader region located 5' of the initiation codon for the initiation codon. The result is a change of DNA base sequence from -TATAAGGATC CCGGGGGCA AGATCTGAGA TATG-Hyg to -TATAAGGATC TGAGATATG-Hyg. The resulting plasmid is known as pCIB3073.

Alternatively, pCIB710 is modified to produce pCIB900, by inserting the Bam HI - Bcl I fragment of pCIB10/35SBt, which contains the 645 amino acid Bt coding sequence, described in Part C4 below, into the Bam HI site of pCIB710 to create pCIB710/35SBt. To introduce an antibiotic resistance marker, pCIB709 is cut with Sal I, a Kpn I/Sal I adaptor is ligated and the resulting ligation product is cut with Kpn I. The Kpn fragment of pCIB709 containing the 35S/hygromycin resistance gene is inserted into the Kpn I site of pCIB710/35SBt to

77

produce pCIB900.

Genes useful as the selectable marker gene include the
hygromycin resistance gene described in Rothstein et al., Gene
53: 153-161 (1987). The hygromycin gene described in this
reference is moved into a pUC plasmid such as pCIB710 or
pCIB709 and the "extra" ATG upstream from the hygromycin
phosphotransferase coding sequence is removed to create
pCIB996. This modified pCIB996 gene is further modified to
remove a BglII, BamHI and SmaI sites from the 5' region of the
gene using standard techniques of molecular biology to make
pCIB3073.

pCIB932 is a pUC19-based plasmid containing the
chimeric gene Pep-C:promoter\Bt\Pep-C:terminator. It is
composed of fragments derived from pPEP-10, a HindIII subclone
of a genomic clone, H1-lambda-14, PNAS USA, 83:2884-2888
(1986), of the maize gene encoding the PEP carboxylase enzyme
active in photosynthesis, and from pCIB930, which is a BamHI
fragment containing the 645 amino acid truncated form of the
the cryIAb endotoxin gene in the BamHI site of pUC18.

The 2.6 kb EcoRI-XhoI fragment from pPEP-10, containing
the polyA addition site from the PEP carboxylase gene, is
isolated and digested with PstI and HincII. The restriction
digest is ligated with PstI/HincII digested pUC18, transformed
into E. coli and transformants screened for those containing a
412 bp PstI-HincII insert in pUC18 and the insert verified by
sequencing. The resulting plasmid is called pCIB931.

The nuclear gene encoding the phosphoenolpyruvate

carboxylase isozyme ("Pep-C") is described in Hudspeth et al.,

Plant Molecular Biology, 12: 579-589 (1989). pCIB932 is

constructed by the ligation of three fragments. The first

fragment, containing the PEP-C transcription terminator, is

produced by digesting pCIB931 to completion with HindIII,

partially with SphI and the 3098 bp fragment isolated. The

second fragment, containing the Bt endotoxin coding sequence,

is produced by digesting pCIB930 with NcoI and SphI and

isolating the 1950 bp fragment. The third fragment, containing

the PEP-C promoter, is produced by digesting pPEP-10 to

completion with HindIII, partially with NcoI and isolating the

2.3 kb fragment. The ligation mix is transformed into E. coli,

transformants with the correct insertion identified and the

insert verified by sequencing.

pCIB932 is cut with PvuII to generate a 4.9 Kb fragment

containing the maize Pep-C:promoter\Bt\Pep-C:terminator and

purified on a 1% LGT agarose gel in 1X TAE. The linearized

pCIB3079 vector and the 4.9 Kb insert from pCIB932 are ligated

using T4 DNA ligase in LGT to make pCIB4401. pCIB4401 is a

maize transformation vector containing the chimeric genes:

35S:promoter\PAT\35S:terminator, Pep-C:promoter\Bt\Pep-C:

terminator, and 35S:promoter\AdhI #1 intron\GUS\35S:

terminator.

Construction of pCIB246 (35S-GUS-35S)

    A CaMV 35S promoter cassette, pCIB246, is constructed

as follows.

    The DdeI restriction site at nucleotide position 7482

of the CaMV genome [Franck et al., Cell, 21:285-294 (1980)] is modified by insertion of a 48 bp oligonucleotide containing several restriction enzyme sites including an NcoI (CCATGG) site, a SalI (GTCGAC) site, and an SstI (GAGCTC) site. This altered CaMV 35S promoter is inserted into a pUC19 vector that had been modified to destroy the vector's SstI and SalI sites. Thus, the CaMV 35S promoter of pCIB1500 contains unique SstI and SalI sites for cloning.

pCIB1500 is digested with SstI/NcoI and ligated with the GUS gene obtained from pBI221 (Clontech Laboratories, Inc., Palo Alto, CA). The NcoI site is fused to the GUS gene such that the ATG of the NcoI site functions as the start codon for the translation of the GUS gene. The CaMV 35S polyadenylation and termination signals are used for the 3' end of the chimeric gene.

Construction of pCIB3069 (35S-Adh1-GUS-35S)

pCIB246 is modified by adding the maize alcohol dehydrogenase gene Adh1 intron number 1 (Adh1) (Dennis et al., Nucleic Acids Research, 12:3983-4000 (1984)) into the Sal I site of pCIB246 to produce plasmid pCIB3007. The Adh1 intron is excised from the maize Adh1 gene as a Bal I/Pst I fragment and subcloned into pUC18 that was cut with Sma I/Pst I to make a plasmid called Adh 1026. Adh 1026 is cut with Pvu II/Sac II, the fragments are made blunt ended with T4 DNA polymerase, Sal I linkers are added using standard procedures and a fragment of about 560 bp is recovered from a 3 % NuSeive gel and ligated into Sal I cut/phosphatase treated pUC18. The Sal I linkered

Adh intron #1 in the resulting plasmid is cut out with Sal I, gel purified, and ligated into Sal I cut/ phosphatase treated pCIB246 to make plasmid pCIB3007.

pCIB3007 is cut with PstI and the ends made blunt by using T4 DNA polymerase (NEW England Biolabs) according to the suppliers' specifications. The resulting blunt ended molecules are cut with Sph I and the approximately 5.8 Kb fragment with one blunt end and one Sph I end is purified on a low gelling temperature (LGT) agarose gel using standard procedures. pCIB900 is cut with Sma I/Sph I and the fragment containing the 35S/Bt gene is purified on a LGT agarose gel. The two gel purified fragments are ligated in LGT agarose using T4 DNA ligase according to standard conditions. The resulting ligated fragments are transformed into E. coli using standard procedures and the resulting plasmid is called pCIB3062. There are two versions of pCIB3062. pCIB3062#1 has a Sma I site regenerated where the Sma I site and the T4 polymerase blunted ends are ligated. This most likely results from the T4 polymerase nibbling a few base pairs from the Pst I site during the blunting reaction. pCIB3062#3 does not have this SmaI site.

pCIB3062#3 is cut with KpnI and made blunt-ended using T4 DNA polymerase, and subsequently cut with Pvu II to yield a 6.4 Kb fragment with blunt ends containing the 35S/GUS and 35S/Bt genes. This blunt-end fragment is ligated into Sma I cut pCIB3073 to produce pCIB3063 or pCIB3069. pCIB3069 contains the same fragment used to make pCIB3063, but the

81

chimeric genes in pCIB3069 are all in the same relative orientation, unlike those in pCIB3063. These plasmids contain a) a 35S promoter operably linked to the hygromycin resistance gene; b) a 35S promoter, with Adh intron #1, operably linked to the GUS gene; and c) a 35S promoter operably linked to a gene coding for the production of the synthetic cryIA(b) insecticidal protein from Bacillus thuringiensis, as described above.

GUS Assays:

GUS assays are done essentially as described in Jefferson, Plant Mol. Bio. Reporter, 5:387-405 (1987). As shown above, plasmid pCIB246 contains a CaMV 35S promoter fused with the GUS gene. The 5' untranslated leader of this chimeric gene contains a copy of the maize Adh1 intron #1. It is used here as a transformation control. Although the same amount of pCIB246 is added to each transformation, the calculated activity varied among Bt constructs tested. The values reported below are averages of 3 replicates. pCIB4407 was tested twice.

pCIB3069   28 nM MU/ug/min

pCIB4407   0.7 nM MU/ug/min, 2.3 nM MU/ug/min

EXAMPLE 5A: ASSAY OF SYNTHETIC cryIA(b) gene FOR INSECTICIDAL
           ACTIVITY AGAINST EUROPEAN CORN BORER

The synthetic cryIA(b) gene in pCIB4414 in E. coli is assayed for insecticidal activity against European corn borer

according to the following protocol.

Molten artifical insect diet is poured into a 60 mm Gellman snap-cap petri dish. After solidification, E. coli cells, suspended in 0.1% Triton X-100, are spread over the surface at a concentration of 3 X 107 cells/cm2. The plates are air dried. Ten first instar European corn borer, Ostrinia nubilalis, which are less than 12 hours old are then placed onto the diet surface. The test is incubated at 30 C in complete darkness for 2-5 days. At the end of the test percent mortality is recorded. A positive clone has been defined as one giving 50% or higher mortality when control E. coli cells give 0-10% background mortality.

For comparison, the native cryIA(b) gene in pCIB3069 is tested at the same concentration. Clones are tested at 3×10^7 cells/cm^2 diet; 20 insects per clone.

The following results are observed:

Clone	Percent Mortality
Control	0
pCIB3069	100
pCIB4414	100

These results indicate that the insecticidal crystal protein produced by the synthetic cryIA(b) gene demonstrates activity against European corn borer comparable to that of the IP produced by the native cryIA(b). Other plasmids containing a synthetic cryIA(b) gene were assayed in a similar manner.

EXAMPLE 5AB: ASSAY OF CRYIA(b) PROTEIN FOR INSECTICIDAL ACTIVITY AGAINST SUGARCANE BORER.

CryIA(b) was expressed in E. coli and assayed for

83

insecticidal activity against Sugarcane borer (<u>Diatrea</u>

<u>saccharalis</u>) according to the same protocol used for European

corn borer, described immediately above. The results are

summarized in the Table.

TABLE

SUGARCANE BORER ASSAY WITH Bt PROTEIN FROM <u>E.</u> <u>COLI</u>

Protein Concentration (ng/g)	Percent Mortality CryIA(b)
10	0
25	0
50	7
100	13
250	40
500	53
1000	80
LC50	380
95% Cl	249-646

The results indicate that the insecticidal protein
produced by a maize optimized Bt gene is effective against
Sugarcane borer. The upper concentrations of CryIA(b) protein,
250 ng/g-1000 ng/g, are achievable in transgenic maize plants
produced in accordance with the instant invention.

EXAMPLE 6: MAIZE PROTOPLAST ISOLATION AND TRANSFORMATION WITH
          THE SYNTHETIC BT GENE

Expression of the synthetic Bt gene is assayed in

transiently transformed maize protoplasts.

Protoplast Isolation Procedure:

1. The contents of 10 two day old maize 2717 Line 6

suspension cultures are pipetted into 50 ml sterile tubes and

allowed to settle. All culture media is then removed and

discarded.

2. Cells (3-5 ml Packed Cell Volume) are resuspended in 30 ml protoplast enzyme solution. Recipe follows:

3% Cellulase RS

1% Macerozyme R10 in KMC Buffer

KMC Buffer (recipe for 1 liter)

KCl	8.65 g
$MgCl_2-6H_2O$	16.47 g
$CaCl_2-2H_2O$	12.50 g
MES	5.0 g

pH 5.6, filter sterilize

3. Mix cells well and aliquot into 100x25 mm petri dishes, about 15 ml per plate. Shake on a gyratory shaker for 4 hours to digest.

4. Pipette 10 ml KMC through each 100 micron sieve to be used. Filter contents of dishes through sieve. Wash sieve with an equal volume KMC.

5. Pipette sieved protoplasts carefully into 50 ml tubes and spin in a Beckman TJ-6 centrifuge for 10 minutes at 1000 rpm (500 x g).

6. Remove supernatant and resuspend pellet carefully in 10 ml KMC. Combine contents of 3 tubes into one and bring volume to 50 ml with KMC.

7. Spin and wash again by repeating the above step.

8. Resuspend all washed protoplasts in 50 ml KMC. Count in a hemocytometer. Spin protoplasts and resuspend at 8 x 10^6/ml in resuspending buffer (RS Buffer).

RS Buffer (recipe for 500 ml)

mannitol   27.33 g

CaCl$_2$ (0.1 M stock) 75 ml

MES   0.5 g

pH 5.8, filter sterilize

Protoplast Transformation Procedure:

1. Aliquot 50 µg plasmid DNA (Bt IP constructs, both synthetic (pCIB4407) and native (pCIB3069)) to 15 ml polystyrene culture tubes. Also aliquot 25 µg GUS-containing plasmid DNA (which does not contain Bt IP (pCIB246) to all tubes. 3 replications are used per construct to be tested, with 1 rep containing no DNA as a control.

Bt constructs: GUS construct:

pCIB3069   pCIB246

pCIB4407

2. Gently mix protoplasts well and aliquot 0.5 ml per tube.

3. Add 0.5 ml PEG-40 to each tube.

PEG-40:

0.4 M mannitol

0.1 M Ca(NO$_3$)$_2$-4H$_2$O

pH 8.0, filter sterilize

4. Mix gently to combine protoplasts with PEG. Wait 30 minutes.

5. Sequentially add 1 ml, 2 ml, and 5 ml W5 solution at 5 minute intervals.

W5 Solution:

86

154 mM NaCl

125 mM $CaCl_2-H_2O$

  5 mM KCl

  5 mM glucose

pH 7.0, filter sterilize

6. Spin for 10 minutes in a Beckman TJ-6 centrifuge at about 1000 rpm (500g). Remove supernatant.

7. Gently resuspend pellet in 1.5 ml FW media and plate carefully in 35x10 mm petri dishes.

FW media (recipe for 1 liter):

MS salts	4.3 g
200X B5 vits.	5 ml
sucrose	30 g
proline	1.5 g
mannitol	54 g
2,4 D	3 mg

pH 5.7, filter sterilize

8. Incubate overnight in the dark at room temperature.

9. Perform GUS assays, insect bioassays, and ELISA's on protoplast extracts as described below.

EXAMPLE 7: CONSTRUCTION OF A FULL-LENGTH SYNTHETIC MAIZE OPTIMIZED CRYIA(b) GENE

Sequence 4 shows the synthetic maize optimized sequence encoding the full-length cryIA(b) insecticidal protein from B. thuringiensis. The truncated version described above represents the first approximately 2 Kb of this gene. The remainder of the full-length gene is cloned using the

87

procedures described above. Briefly, this procedure entails synthesizing DNA oligomers of 40 to 90 NT in length, typically using 80 mers as an average size. The oligomers are purified using standard procedures of HPLC or recovery from a polyacrylamide gel. Purified oligomers are kinased and hybridized to form fragments of about 500 bp. The hybridized oligomers can be amplified using PCR under standard conditions. The 500 bp fragments, either directly from hybridizations, from PCR amplification, or recovered from agarose gels after either hybridization or PCR amplification, are then cloned into a plasmid and transformed into E. coli using standard procedures.

Recombinant plasmids containing the desired inserts are identified, as described above, using PCR and/or standard miniscreen procedures. Inserts that appear correct based upon their PCR and/or restriction enzyme profile are then sequenced to identify those clones containing the desired open reading frame. The fragments are then ligated together with the approximately 2 Kb synthetic sequence described in Example 2 to produce a full-length maize optimized synthetic cryIA(b) gene useful for expression of high levels of CryIA(b) protein in maize.

G+C Content of native and synthetic Bt genes:

Full-length native	38.8%
Truncated native	37.2%
Full-length synthetic	64.8%
Truncated synthetic	64.6%

% homology of the final truncated version of the Bt gene

relative to a "pure" maize codon usage gene: 98.25%

EXAMPLE 8 Construction of a plant expressible, full-length,

hybrid partially maize optimized cryIA(b) gene.

pCIB4434 contains a full length CryIA(b) gene

comprised of about 2 Kb of the synthetic maize optimized

cryIA(b) gene with the remainder (COOH terminal encoding

portion) of the gene derived from the native gene. Thus, the

coding region is a chimera between the synthetic gene and the

native gene, but the resulting protein is identical to the

native cryIA(b) protein. The synthetic region is from

nucleotide 1-1938 (amino acids 1 to 646) and the native coding

sequence is from nucleotide 1939-3468 (amino acids 647 to

1155). The sequence of this gene is set forth in Fig. 7. A

map of pCIB4434 is shown in Fig. 8.

The following oligos were designed to make pCIB4434:

KE134A28 = 5'-CGTGACCGAC TACCACATCG ATCAAGTATC CAATTTAGTT

GAGT-3'

KE135A28 = 5'-ACTCAACTAA ATTGGATACT TGATCGATGT GGTAGTCGGTC

ACG-3'

KE136A28 = 5'-GCAGATCTGA GCTCTTAGGT ACCCAATAGC GTAACGT-3'

KE137A28 = 5'-GCTGATTATG CATCAGCCTAT-3'

89

KE138A28 = 5'-GCAGATCTGA GCTCTTATTC CTCCATAAGA AGTAATTC-3'

MK05A28 = 5'-CAAAGGTACC CAATAGCGTA ACG-3'

MK35A28 = 5'-AACGAGGTGT ACATCGACCG-3'

pCIB4434 is made using a four-way ligation with a 5.7 kb fragment from pCIB4418, a 346 bp Bst E II\Kpn I PCR-generated synthetic:native fusion fragment, a 108 bp Kpn I\Nsi I native CryIA(b) fragment from pCIB1315, and a 224 bp Nsi I\Bgl II PCR-generated fragment. Standard conditions for ligation and transformation are as described previously.

A synthetic:native gene fusion fragment is made in two steps using PCR. The first 253 bp of the PCR fusion fragment is made using 100 pmols of oligos KE134A28 and MK04A28 with approximately 200 ng of native cryIA(b) template in a 100 ul volume with 200 nm of each dNTP, 1 X PCR buffer (Perkin Elmer Cetus), 20 % glycerol, and 5 units of Taq polymerase (Perkin Elmer Cetus). The PCR reaction is run with the following parameters: 1 minute at 94°C, 1 minute at 55°C, 45 seconds at 72°C, with extension 3 for 3 seconds for 25 cycles. A fraction (1 %) of this first PCR reaction is used as a template along with 200 ng of the synthetic cryIA(b) DNA to make the complete 351 bp synthetic:native fusion fragment. Oligos used as PCR primers in this second PCR reaction are 50 pmols of MK35A28, 50 pmols of MK04A28, and 25 pmols of KE135A28. The PCR reaction mix and parameters are the same as those listed above. The

resultant 351 bp synthetic:native fusion fragment is treated

with Proteinase K at 50 ug\ml total concentration and

phenol\chloroform extraction followed by ethanol precipitation

before cutting with Bst E II\Kpn I using standard conditions.

The 224 bp Nsi I\Bgl II PCR fragment used in making

pCIB4434 is made using 100 pmols of oligos KE137A28 and

KE138A28 and 200 ng of the native cryIA(b) gene as template in

100 ul volume with the same PCR reaction mix and parameters as

listed above. The 230 bp PCR native cryIA(b) fragment is

treated with Proteinase K, phenol\chloroform extracted, and

ethanol precipitated as described above, before cutting with

Nsi I\Bgl II.

pCIB4434 was transformed into maize protoplasts as

described above. Line 6 2717 protoplasts were used with

pCIB4434 and pCIB4419 as a control for comparison. The results

are shown below:

                              ng Bt/mg protein

4419(35S)   14,400 ± 2,100

4434(full-length)   2,200 ± 900

Background = 13 ng Bt/mg protein for untransformed protoplasts

The results indicate that pCIB4434 expresses at a level

91

of about 15% of pCIB4419.

Western blot analysis shows at least one-third of the cryIA(b) protein produced by pCIB4434 in this system is about 130 kD in size. Therefore, a significant amount of full-length cryIA(b) protein is produced in maize cells from the expression of pCIB4434.

EXAMPLE 7. Construction of a full-length, cryIA(b) genes encoding a temperature-stable cryIA(b) protein.

Constructs pCIB5511-5515, each containing a full-length, cryIA(b) gene are described below. In these sequences, the 26 amino acid deletion between amino acids 793 and 794, KCGEPNRCAPHLEWNPDLDCSCRDGE, present in cryIA(a) and cryIA(c) but not in cryIA(b), has been repaired. The gene in pCIB5513 is synthetic; the other four genes are hybrids, and thus are partially maize optimized.

Construction of pCIB5511

This plasmid is a derivative of pCIB4434. A map of pCIB5511 is shown in Fig. 10. A 435 bp segment of DNA between bp 2165 and 2590 was constructed by hybridization of synthetic oligomers designed to represent the upper and lower strand as described above for the construction of the truncated cryIA(b) gene. This segment of synthetic DNA is synthesized using standard techniques known in the art and includes the 26 amino acid deletion found to occur naturally in the cryIA(b) protein in _Bacillus thuringiensis_ kurstaki HD-1. The entire inserted segment of DNA uses maize optimized codon preferences to encode amino acids. The 26 amino acids used to repair the naturally

92

occurring deletion are contained within this fragment. They are
inserted starting at position 2387 between the KpnI site at nt
2170 and the XbaI site at nt 2508 (2586 in pCIB5511) of
pCIB4434. pCIB5511 is constructed via a three way ligation
using a 3.2 Kb fragment obtained by restriction digestion of
pCIB4434 with SphI and KpnI, a 3.8 Kb fragment obtained by
digestion of pCIB4434 with SphI and XbaI, and a 416 bp fragment
obtained by digestion of the synthetic DNA described above,
with KpnI and XbaI. Enzymatic reactions are carried out under
standard conditions. After ligation, the DNA mixture is
transformed into competent _E. coli_ cells using standard
procedures. Transformants are selected on L-agar containing 100
µg/ml ampicillin. Plasmids in transformants are characterized
using standard mini-screen procedures. The sequence of the
repaired cryIA(b) gene encoding the cryIA(b) temperature (heat)
stable protein is set forth in Fig. 9.

<div style="text-align:center">Construction of pCIB5512</div>

This plasmid construct is a derivative of pCIB4434.
A map of pCIB5512 is shown in Fig. 12. DNA to repair the 26
amino acid deletion is prepared using standard techniques of
DNA synthesis and enzymatic reaction. Three double stranded DNA
cassettes, pGFcas1, pGFcas2 and pGFcas3, each about 300 bp in
size, are prepared. These cassettes are designed to contain the
maize optimized codons while maintaining 100% amino acid
identity with the insecticidal protein. These cassettes are
used to replace the region between restriction site BstEII at
position 1824 and XbaI at position 2508 and include the

<div style="text-align:center">93</div>

insertion of the additional 78 bp which encode the missing 26 amino acids (described above for pCIB5511 in pCIB4434). Each of these cassettes is cloned into the EcoRV site of the vector Bluescript (Stratagene) by standard techniques. The three cassettes are designed to contain overlapping restriction sites. Cassette 1 has restriction sites BstEII at the 5' end and EcoRV at the 3' end: cassette 2 has EcoRV at the 5' end and ClaI at the 3' end and cassette 3 has ClaI at the 5' end and Xba I at the 3' end. They are cloned individually in Bluescript and the the complete 762 bp fragment is subsequently assembled by ligation using standard techniques. pCIB5512 is assembled using this 762 bp fragment and ligating it with a 6.65 Kb fragment obtained by a complete digestion of pCIB4434 with BstEII and a partial digestion with XbaI. Alternatively, a four way ligation using the same vector and the three cassettes digested with the specific enzymes can be employed. Enzymatic reactions are carried out under standard conditions. After ligation, the DNA mixture is transformed into competent E. coli cells using standard procedures. Transformants are selected on L-agar containing 100 µg/ml ampicillin. Plasmids in transformants are characterized using standard mini-screen procedures. The resulting plasmid is pCIB5512. The sequence of the repaired cryIA(b) gene is illustrated in Fig. 11. This repaired cryIA(b) differs from that carried in pCIB5511 in that a larger region of the cryIA(b) coding region is optimized for maize expression by using maize preferred codons.

Construction of pCIB5513

94

This plasmid contains a repaired cryIA(b) gene derived
from pCIB5512. A map of pCIB5513 is shown in Fig. 14. The
region 3' from the XbaI site at position 2586 to the end of the
gene (BglII site at position 3572) is replaced entirely with
maize optimized codons. This region is synthesized, using
standard techniques of DNA synthesis and enzymatic reaction,
well known in the art, as four double stranded DNA cassettes
(cassettes # 4,5 ,6 ,7). Adjacent cassettes have overlapping
restriction sites to facilitate assembly between cassettes.
These are XbaI and XhoI at the 5' and 3' ends of cassette 4;
XhoI and SacI at the 5' and 3' ends, respectively, of cassette
5; SacI and BstXI at the 5' and 3' ends, respectively, of
cassette 6; and BstXI and BglII at the 5' and 3' ends,
respectively, of cassette 7. As described for pCIB5512, the
cassettes are cloned into the blunt-end EcoRV site of the
Bluescript vector (Stratagene) and the full-length "repaired"
cryIA(b) gene cloned either by sequential assembly of the above
cassettes in Bluescript followed by ligation of the complete
967 bp synthetic region with a 6448 bp fragment obtained by a
complete digestion of pCIB5512 with BglII and a partial
digestion with XbaI. Alternately, the plasmid containing the
full-length genes is obtained by a 5-way ligation of each of
the four cassettes (after cleavage with the appropriate
enzymes) and the same vector as above. The sequence of the
full-length, "repaired" cryIA(b) gene is set forth in Fig. 13.
The protein encoded by the various synthetic and
synthetic/native coding region chimeras encode the same

protein. This protein is the heat-stable version of cryIA(b) produced by repairing the naturally occurring 26 amino acid deletion found in the cryIA(b) gene from <u>Bacillus</u> <u>thuringiensis</u> kurstaki HD-1 when the homologous region is compared with either cryIA(a) or cryIA(c) <u>Bacillus</u> <u>thuringiensis</u> delta-endotoxins.

<div align="center">Construction of pCIB5514</div>

This plasmid is a derivative of pCIB4434. A map of pCIB5514 is shown in Fig. 16. It is made using synthetic DNA cassette #3 (see above) which contains a maize optimized sequence of the region between the ClaI site (position 2396) found in the 26 amino acid thermostable region and the XbaI site at position 2508 in pCIB4434 (2586 in pCIB5511). The region between nt 2113 of pCIB4434 and the junction of the thermostable region is PCR amplified by using pCIB4434 as template with the following primers:

forward: 5'GCACCGATATCACCATCCAAGGAGGCGATGACGTATTCAAAG-3'

reverse:

5'-AGCGCATCGATTCGGCTCCCCGCACTTGCCGATTGGACTTGGGGCTGAAAG-3'.

The PCR product is then digested with restriction enzymes KpnI and ClaI and ligated in a four part reaction with a 189 bp fragment obtained by digestion of cassette 3 with ClaI and XbaI, a 3.2 Kb fragment of pCIB4434 digested with SphI and KpnI, and a 3.8 Kb fragment of pCIB4434 obtained by digestion with SphI and Xba. Enzymatic reactions are carried out under standard conditions. The ligation product is transformed into competent <u>E</u> <u>coli</u> cells, selected with ampicillin and screened

using standard procedures described above. The sequence of the repaired cryIA(b) gene contained in pCIB5514 is shown in Fig. 15.

<div align="center">Construction of pCIB5515</div>

pCIB4434 was modified by adding the 78bp Geiser thermostable element (Geiser TSE), described above, between the Kpn I site (2170 bp) and the Xba I site (2508 bp) in the native Btk region. The exact insertion site starts at the nucleotide #2379. The region containing the Geiser TSE was amplified by two sets of PCR reactions, i.e. the Kpn I - Geiser TSE fragment and the Geiser TSE - Xba I fragment.

PCR primer#1:   (Kpn I site)

        5' - ATTACGTTAC GCTATTGGGT ACCTTTGATG - 3'

PCR primer#2:   (Geiser TSE bottom)

        5' - TCCCCGTCCC TGCAGCTGCA GTCTAGGTCC GGGTTCCACT

                  CCAGGTGCGG AGCGCATCGA TTCGGCTCCC CGCACTTGCC

                  GATTGGACTT GGGGCTGA - 3'

PCR primer#3:   (Geiser TSE top)

        5' - CAAGTGCGGG GAGCCGAATC GATGCGCTCC GCACCTGGAG

                  TGGAACCCGG ACCTAGACTG CAGCTGCAGG GACGGGGAAA

                  AATGTGCCCA TCATTCCC - 3'

PCR primer#4:   (Xba I site)

        5' - TGGTTTCTCT TCGAGAAATT CTAGATTTCC - 3'

<div align="center">97</div>

After the amplification, the PCR fragments were digested with (Kpn I + Cla I) and (Cla I + Xba I), respectively. These two fragments were ligated to the Kpn I and Xba I digested pCIB4434. The resulting construct pCIB5515 is pCIB4434 with a Geiser TSE and an extra Cla I site flanked by Kpn I and Xba I. A map of pCIB5515 is illustrated in Fig. 38. The cryIA(b) gene contained herein, which encodes a temperature stable cryIA(b) protein, is shown in Fig. 37.

Examples 9-20 set forth below are directed to the isolation and characterization of a pith-preferred promoter.

EXAMPLE 9.   RNA Isolation and Northern Blots

All RNA was isolated from plants grown under greenhouse conditions. Total RNA was isolated as described in Kramer et al., Plant Physiol., 90:1214-1220 (1990) from the following tissues of Funk maize line 5N984: 8, 11, 15, 25, 35, 40, and 60 day old green leaves; 8, 11, 15, 25, 35, 39, 46, 60 and 70 day old pith; 60 and 70 day old brace roots from Funk maize line 5N984; 60 and 70 day 5N984 sheath and ear stock. RNA was also isolated from 14 day 211D roots and from developing seed at weekly intervals for weeks one through five post-pollenation. Poly A+ RNA was isolated using oligo-dT as described by Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd ed.), 1989, and Northern blots were carried out, also as per Sambrook et al. using either total RNA (30 µg) or poly A+ RNA (2-10 µg). After electrophoresis, RNA was blotted onto

Nitroplus 2000 membranes (Micron Separations Inc). The RNA was linked to the filter using the Stratalinker (Stratagene) at 0.2 mJoules. The northerns were probed with the 1200 bp EcoRI pith (TRpA) 8-2 cDNA fragment, isolated by using 0.8% low melting temperature agarose in a TBE buffer system. Northerns were hybridized and washed and the filters exposed to film as described in Isolation of cDNA clones.

EXAMPLE 10. Isolation of cDNA Clones

First strand cDNA synthesis was carried out using the BRL AMV reverse transcriptase system I using conditions specified by the supplier (Life Technologies, Inc., Gaithersburg, MD). Specifically, 25 μl reactions containing 50 mM Tris-HCl pH 8.3, 20 mM KCl, 1 mM DTT, 6 mM MgCl2, 1 mM each of each dNTP, 0.1 mM oligo (dT)12-18, 2 μg pith poly(A+) RNA, 100 μg/ml BSA, 50 μg/ml actinomycin D, 8 units placental RNase inhibitor, 1 μl (10 mM Ci/ml) 32P dCTP >3000 mCi/mM as tracer, and 30 units AMV reverse transcriptase were incubated at 42°C for 30 min. Additional KCl was added to a concentration of 50 mM and incubation continued a further 30 min. at 42°C. KCl was added again to yield a final concentration of 100 mM. Additional AMV reverse transcriptase reaction buffer was added to maintain starting concentrations of the other components plus an additional 10 units, and the incubation continued at 42°C for another 30 min. Second strand synthesis was completed using the Riboclone cDNA synthesis system with Eco RI linkers (Promega, Madison, WI). Double stranded cDNA was sized on an

1% agarose gel using Tris-borate-EDTA buffer as disclosed in
Sambrook et al., and showed an average size of about 1.2 Kb.
The cDNA was size fractionated using NA45 DEAE membrane so as
to retain those molecules of about 1000 bp or larger using
conditions specified by the supplier (Schleicher and Schuell).
Size fractionated cDNA was ligated into the Lambda ZapII vector
(Stratagene, La Jolla, CA) and packaged into lambda particles
using Gigapack II Plus (Stratagene, La Jolla, CA). The
unamplified library had a titer of 315,000 pfu while the
amplified library had a titer of 3.5 billion/ml using PLK-F'
cells.

Recombinant phage were plated at a density of 5000 pfu
on 150 X 15mm L-agar plates. A total of 50,000 phage were
screened using duplicate lifts from each plate and probes of
first strand cDNA generated from either pith derived mRNA or
seed derived mRNA. The lifts were done as described in
Sambrook et al. using nitrocellulose filters. DNA was fixed to
the filters by UV crosslinking using a Stratalinker
(Stratagene, La Jolla, CA) at 0.2 mJoule. Prehybridization and
hybridization of the filter were carried out in a solution of
10X Denhardts solution, 150 µg/ml sheared salmon sperm DNA, 1%
SDS, 50 mM sodium phosphate pH 7, 5 mM EDTA, 6X SSC, 0.05%
sodium pyrophosphate. Prehybridization was at 62°C for 4 hours
and hybridization was at 62°C for 18 hours (overnight) with 1
million cpm/ml in a volume of 40 ml. Filters were washed in
500 ml of 2X SSC, 0.5% SDS at room temperature for 15 min. then
at 63°C in 0.1X SSC, 0.5% SDS for 30 min. for each wash.

100

Radiolabeled DNA probes were made using a BRL random prime
labeling system and unincorporated counts removed using Nick
Columns (Pharmacia). Filters were exposed overnight to Kodak
X-Omat AR X-ray film with (DuPont) Cronex Lightning Plus
intensifying screens at -80°C. Plaques showing hybridization
with the pith-derived probe and not the seed-derived probe were
plaque purified for further characterization.

EXAMPLE 11. Isolation of Genomic Clones

        Genomic DNA from Funk inbred maize line 211D was
isolated as described by Shure et al., Cell, 35:225-233 (1988).
The DNA was partially digested with Sau 3A and subsequently
size fractionated on 10-40% sucrose gradients centrifuged in a
Beckman SW40 rotor at 22,000 rpm for 20 hours at 20°C.
Fractions in the range of 9-23 Kb were pooled and ethanol
precipitated. Lambda Dash II (Stratagene) cut with Bam HI was
used as described by the supplier. The library was screened
unamplified and a total of 300,000 pfu were screened using the
conditions described above. The library was probed using pith-
specific (TrpA) cDNA clone 8-2, pCIB5600 which was identified
in the differential screen of the cDNA library. Isolated
clones were plaque purified and a large scale phage preparation
was made using Lambdasorb (Promega) as described by the
supplier. Isolated genomic clones were digested with Eco RI
and the 4.8 kb EcoRI fragment was subcloned into Bluescript
vector (Stratagene).

                            101

EXAMPLE 12. DNA Sequence and Computer Analysis

Nucleotide sequencing was performed using the dideoxy chain-termination method disclosed in Sanger et al., PNAS, 74:5463-5467 (1977). Sequencing primers were synthesized on an Applied Biosystems model 380B DNA synthesizer using standard conditions. Sequencing reactions were carried out using the Sequenase system (US Biochemical Corp.). Gel analysis was performed on 40 cm gels of 6% polyacrylamide with 7 M urea in Tris-Borate-EDTA buffer (BRL Gel-Mix 6). Analysis of sequences and comparison with sequences in GenBank were done using the U. of Wisconsin Genetic Computer Group Sequence Analysis Software (UWGCG).

EXAMPLE 13. Mapping the Transcriptional Start Site

Primer extension was carried according to the procedure of Metraux et al., PNAS, 86:896-900 (1988). Briefly, 30 μg of maize pith total RNA were annealed with the primer in 50 mM Tris pH 7.5, 40 mM KCl, 3 mM MgCl2 (RT buffer) by heating to 80°C for 10 minutes and slow cooling to 42°C. The RNA/primer mix was allowed to hybridize overnight. Additional RT buffer, DTT to 6 mM, BSA to 0.1 mg/ml, RNAsin at 4 U/ml and dNTP's at 1 mM each were added. Then 8 units AMV reverse transcriptase were added and reaction placed at 37°C for one hour. The primer used was 5'-CCGTTCGTTC CTCCTTCGTC GAGG-3', which starts at +90 bp relative to the transcription start. See Fig. 29A. A sequencing ladder using the same primer as in the primer extension reaction was generated using the 4.8 Kb genomic clone

to allow determination of the transcriptional start site. The

sequencing reaction was carried out as described in Example 12.

        RNase protection was used to determine if the the 371

bp sequence from +2 bp to +373 bp (start of cDNA) was

contiguous or if it contained one or more introns. A 385 bp

SphI-NcoI fragment spanning +2 bp to +387 bp relative to

transcriptional start see Fig. 29B was cloned into pGEM-5Zf(+)

(Promega) and transcribed using the Riboprobe Gemini system

(Promega) from the SP6 promoter to generate radioactive

antisense RNA probes as described by the supplier. RNase

protection was carried out as described in Sambrook et al.

pBR322 (cut with HpaII and end labelled with 32P-dCTP) and

Klenow fragment were used molecular weight markers. Gels were

6% acrylamide/7M urea (BRL Gel-Mix 6) and were run at 60 watts

constant power.

EXAMPLE 14. Genomic Southern Blots

        Genomic DNA was isolated from maize line 211D using the

procedure of Shure et al., supra. 8 μg of genomic DNA were

used for each restriction enzyme digest. The following enzymes

were used in the buffer suggested by the supplier: BamHI,

EcoRI, EcoRV, HindIII, and SacI. Pith cDNA clone number 8-2

was used for estimating gene copy number. The digested DNA was

run on a 0.7% agarose gel using Tris-Borate-EDTA buffer system.

The gel was pretreated with 250 mM HCl for 15 min. to

facilitate transfer of high molecular weight DNA. The DNA was

transferred to Nitroplus 2000 membrane and subsequently probed

with the pith cDNA 8-2. The blot was washed as described
in Example 10.

<u>EXAMPLE 15.</u> PCR Material and Methods

PCR reactions were preformed using the GeneAmp DNA
Amplication reagent kit and AmpliTaq recombinant Taq DNA
polmerase (Perkin Elmer Cetus). Reaction condition were as
follows: 0.1 to 0.5 uM of each of the two primers used per
reaction, 25 ng of the pith 4.8 Kb EcoRI fragment in
Bluescript, plus the PCR reaction mix described by the supplier
for a total volume of 50 uL in 0.5 mL GeneAmp reaction tube
(Perkin Elmer Cetus). The DNA Thermal Cycler (Perkin Elmer
Cetus) using the Step-Cycle program set to denature at 94°C for
60 s, anneal at 55°C for 60 s, and extend at 72°C for 45 s
followed by a 3-s-per-cycle extension for a total of 30 cycles.
The following primer sets were used: I. 83 X 84, -429 bp to -2
bp; II. 49 x 73, -69 bp to +91 bp; III. 38 X 41, +136 bp to
+258 bp; and IV. 40 X 75, +239 bp to +372 bp. These are marked
on Fig. 24.

<u>EXAMPLE 16.</u> Isolation of a Pith-Preferred Gene.

A cDNA library derived from pith mRNA cloned into
Lambda Zap and screened using first strand cDNA derived from
either pith or seed mRNA. Clones which hybridized with only
the pith probe were plaque purified and again screened. Clones
passing the second screen were used as probes in northern blots
containing RNA from various maize tissues.

EXAMPLE 17. Gene Structure and Sequence Analysis.

The 1.2 Kb insert of the cDNA clone 8-2 was sequenced using the dideoxy method of Sanger et al., supra. Likewise, the genomic equivalent contained on a 4.8 Kb EcoRI fragment in Bluescript denoted as pCIB5601, was sequenced. This information revealed that the genomic copy of the coding region spans 1.7 Kb and contains five introns. The mRNA transcript represents six exons. This is shown in Fig. 24. The exons range in size from 43 bp to 313 bp and the introns vary in size from 76 bp to 130 bp. The entire sequence of the gene and its corresponding deduced amino acid sequence are shown in Fig. 24.

This gene encodes a protein of 346 amino acids with a molecular mass of about 38 kD. As illustrated in Table 1, the predicted protein shows 62% similarity and 41% identity with the subunit protein of Pseudomonas aeruginosa and has high homology with trpA proteins from other organisms.

Table 1

Conservation of TrpA sequences between a maize TrpA gene and other organisms.

Organisms compared	% amino acid Similarity	% amino acid Identity
Haloferax volancii	56.4	36.1
Methanococcus voltae	58.1	35.1
Pseudomonas aeruginosa	62.5	41.8
Neurospora crassa	61.4	39.3
Saccharomyces cerevisiae	56.7	36.1

Similarity groupings, I=L=M=V, D=E, F=Y, K=R, N=Q, S=T

Similarities and indentities were done using the GAP program
from UWGCG.
              Crawford et al., <u>Ann. Rev. Microbiol.</u>, 43:567-600

(1989), incorporated herein by reference, found regions of

conserved amino acids in bacterial trpA genes. These are amino

acids 49 to 58, amino acids 181 to 184, and amino acids 213 to

216, with the rest of the gene showing greater variability than

is seen in the TrpB sequence. An alignment of known trpA

proteins with the maize TrpA protein (not shown) illustrates

that the homology between the maize gene and other trpA

proteins is considerable. Also, it is comparable to the level

of homology observed when other TrpA proteins are compared to

each other as described in Crawford et al., supra.

          To determine the location of the transcription start

site and whether or not there were introns present in this

region, four polymerase chain reaction (PCR) generated

fragments of about 122 bp to 427 bp from the region -429 bp to

+372 bp were used for northern analysis. The results of the

northerns showed that PCR probes II, III, IV hybridized to pith

total RNA and PCR probe I did not hybridize. This indicated

that the transcription start was in the -69 bp to +90 bp

region. To more precisely locate the transcriptional start

site, primer extension was employed. Fig. 29A shows that when

a primer (#73) located at +90 bp relative to the

transcriptional start is used for primer extension, the

transcriptional start site is located at +1, 1726 bp on the

genomic sequence.

The first ATG from the transcriptional start site is at +114 bp. This is the ATG that would be expected to serve as the site for translational initiation. This ATG begins an open reading that runs into the open reading frame found in the cDNA clone. The first 60 amino acids of this predicted open reading frame strongly resemble a chloroplast transit peptide. See Berlyn et al. PNAS, 86:4604-4608 (1989) and Neumann-Karlin et al., EMBO J., 5:9-13 (1986). This result suggests that this protein is targeted to a plastid and is likely processed to yield the active protein. Transient expression assays in a maize mesophyll protoplast system using a maize optimized B.t. gene driven by the trpA promoter showed that when the ATG at +114 bp is used as the fusion point, the highest levels of expression are obtained. Using either of the next two ATGs in the sequence substantially reduces the level of expression of the reporter gene. The ATG at +390 bp gave some activity, but at a much lower level than the +114 ATG, and the ATG at +201 bp gave no activity.

Athough a number of TATA like boxes are located upstream of the upstream of the transcriptional start site at +1 bp, the TATAAT at -132 bp is most like the plant consensus of TATAAA. See Joshi, Nuc. Acids Res., 15:6643-6653 (1987). The presumptive CCAAT like box was found at -231 bp . The nucleotide sequence surrounding the ATG start (GCGACATGGC) has homology to other maize translation starts as described in Messing et al., Genetic Engineering of Plants: An Agricultural

107

Perspective, Plenum Press, pp. 211-227 (1983), but differs from that considered a consensus sequence in plants (ANNATGGC). *See*, *Joshi*, above. The presumptive poly(A) addition signal is located at 3719 bp (AATAAA) on the genomic sequence, 52 bp from the end of the cDNA. The sequence matches known sequences for maize as described in Dean et al., *Nuc. Acids Res.*, 14:2229-2240 (1986), and is located 346 bp downstream from the end of protein translation. *See* Dean et al., *Nuc. Acids Res.*, 14:2229-2240 (1986). The 3' untranslated sequence of the cDNA ends at 3775 bp on the genomic sequence.

Fig. 28 shows a Southern blot of maize 211D genomic DNA with the approximate gene copy number as reconstructed using pith gene 8-2 cDNA. From the restriction digests and reconstruction there appear to be 1-2 copies of the gene present per haploid genome. There do not appear to be other genes with lower levels of homology with this gene. Therefore, this represents a unique or small member gene family in maize.

EXAMPLE 18. RNase Protection

The structure of the 5' end of the mRNA was determined using RNase protection. The RNase protection was carried out using a probe representing 385 nt from +2 bp to +387 bp. This region from the genomic clone was placed in the RNA transcription vector pGEM-5Zf(+) and a 32P labelled RNA probe generated using SP6 polymerase. The probe and the extra bases from the multiple cloning site produce a transcript of 461 nt. The probe was hybridized with total pith RNA and subsequently

108

digested with a mixture of RNase A and T1 and the protected

fragments analyzed on denaturing polyacrylamide gels. Analysis

of the gels shows a protected fragment of about 355 nt and

another fragment of about 160 nt. See Fig. 29B.

The fact that primer extension using a primer (#73) at

+80 bp produces a product of 90 NT in length argues that the

5' end of the transcript is located at position +1 bp. Primer

extension from a primer in this region produces a product, so

one would expect this also to be detected by the RNase

protection assay. This primer is located in the 5' region of

the RNase protection probe. The cDNA clone contains sequences

present in the 3' end of the RNase protection probe and hence

were expected to be protected in this assay. Since only one

band is present on the gel which could account for both of

these sequences, we are confident that the protected fragment

is indeed the larger band and that the smaller single band is

an artifact. If there were an intron in this region, fragments

from each end would be present in the probe, and hence would be

detectable on the gel. Of the two bands seen, one of them

appears to represent the entire 5' region, therefore we do not

believe that there is an intron located in this region.

EXAMPLE 19. Complementation of E. coli TrpA Mutant with the

Pith cDNA 8-2

E. coli strain CGSC strain 5531 from the E. coli

Genetic Stock Center, Yale University (O.H. Smith lab strain

designation, #M5004) with chromosomal markers glnA3, TrpA9825,

1-, IN(rrnD-rrnE), thi-1 as described in Mayer et al., Mol. Gen.
Gentet., 137:131-142 (1975), was transformed with either the
pith (TRpA) cDNA 8-2 or Bluescript plasmid (Stratagene) as
described in Sambrook et al., supra. The transformants
containing the TrpA cDNA 8-2 had the ability to grow without
the presence of tryptophan on minimal medium whereas the
transformants with the Bluescript (Stratagene) plasmid or
untransformed control were not able to grow without tryptophan.
The cells transformed with the maize TrpA gene grew very slowly
with colonies visible after seven days growth at room
temperature. All strains were grown on M9 minimal medium
supplemented with 200 ug/ml glutamine, 0.01 ug/ml thiamine and
with or without 20 ug/ml tryptophan. All transformants were
checked for the presence of the appropriate plasmid by
restriction enzyme analysis. Colonies growing in the absence
of tryptophan all contained clone 8-2 containing the cDNA for
the putative maize TrpA gene, as confirmed by Southern
hybridization (data not shown). These results support the
conclusion that this is the maize tryptophan synthase subunit A
protein.

EXAMPLE 20. Gene Expression

        The expression pattern of the pith-preferential gene
throughout the plant was examined. Different maize genotypes
were also examined for patterns of expression of this gene.
The following tissues were used as the source of RNA for these
studies: upper, middle, and lower pith, brace roots, ear
shank, cob in genotype 5N984; upper, middle, lower pith, 10 day

old leaves, 14 day old roots and pith from the entire plant in

genotype 211D, and seed from genotype 211D which had been

harvested at weekly intervals one to five weeks

post-pollination. Lower pith is derived from, i.e. constitutes

the two internodes above brace roots; middle pith is derived

from the next three internodes; upper pith represents the last

two internodes before the tassel in 60 and 70 day plants. Only

two internodes were present in 39 day old plants and three

internodes for 46 day old plants. Northern blot analysis shows

that transcripts hybridizing with a probe derived from the pith

cDNA accumulate rapidly in young pith and young leaf. As the

age of the plant increases and one moves up the stalk, there is

a significant decrease in the amount of transcript detected.

See Figs. 25A-D. At no time is message from this gene detected

in seed derived RNA, either total RNA or poly A+ RNA. See Fig.

26. Transcript is also detected in root, earshank, and sheath

but not at the high levels detected in the pith and young leaf

tissues. See Figs. 25B, 25C. Some message is detected in

brace roots, but only at a very low level. See Fig. 25D. Six

maize undifferentiated callus lines were analyzed by northern

blot analysis and no expression was found for this gene (data

not shown) in any callus sample. The level of expression of

this gene is extremely high since a very strong signal to a

probe from TrpA gene 8-2 can be detected in pith and leaf as

little as two hours after exposure of the blot to film (Fig.

25A). The amount of mRNA made is comparable to that derived

from the maize phosphoenolpyruvate carboxylase gene disclosed

111

in Hudspeth et al., Plant Mol. Biology, 12:579-589 (1989),
another highly expressed maize gene. Hudspeth is incorporated
herein by reference.

The expression pattern of this gene is not temporally
constant. Expression is very high in the lower and middle pith
of plants less than 60 days old and decreases rapidly near the
top of the plant. As the plant reaches maturity, e.g. over 70
days old, the expression drops to nearly undetectable levels
except in the lower pith and earshank. The accumulation of
transcript in young leaf is nearly as high as that seen in
lower pith but expression decreases rapidly and is undetectable
in leaves over 40 days of age. Expression in leaf was found to
be variable depending on the season when it is grown.

Examples 21-39 set forth below are directed to the
isolation, characterization and expression analysis of a
pollen-specific promoter according to the present invention.

Identification of pollen-specific proteins

Example 21. Maize Plant Growth

Maize plants (Zea mays Funk inbred 211D) were grown
from seed in a vermiculite/sand mixture in a greenhouse under a
16 hour light/8 hour dark regime.

Example 22. Total Pollen Protein Isolation

Mature pollen was isolated from maize plants at the
time of maximum pollen shed. It was sieved to remove debris,
frozen in liquid nitrogen, and a 3-4 ml volume of frozen pollen
was ground in a mortar and pestle with an equal volume of
75-150 μm glass beads. 40 ml of grinding buffer (2mM EDTA, 5mM

DTT, 0.1% SDS, 100 mM Hepes pH 8) was added and the mixture was

ground again. The glass beads and intact pollen grains were

pelleted by low speed centrifugation, and mixture was clarified

by centrifugation at 10,000 g for 15 minutes. Protein was

precipitated from the supernatant by addition of acetone to

90%.

Example 23. Pollen Exine Protein Isolation

    Exine Protein was isolated from maize 211D shed pollen

as described in Matousek and Tupy, J., _Plant Physiology_

119:169-178 (1985).

Example 24. Leaf Protein Isolation

    Young leaves (about 60% expanded) were cut from the

maize plant the midrib removed. Total protein was isolated as

for pollen, except that the material was not frozen and

grinding was in a Waring blender without glass beads.

Example 25. Kernel Protein Isolation

    Ears with fully developed, but still moist kernels were

removed from the plant and the kernels cut off with a scalpel.

Total protein was isolated as for leaves.

Example 26. Gel Electrophoresis of Maize Proteins

    Pollen, leaf and kernel proteins were separated on SDS

polyacrylamide gels as described in Sambrook et al, _Molecular_

Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory

Press: New York (1989). Following staining by Coomasie blue,

protein bands from pollen, leaf and kernel were compared and

abundant proteins of approximately 10 kD, 13 kD, 20 kD, 45 kD,

55 kD and 57 kD were determined to be pollen specific.

Identification of Pollen-Specific cDNA clones

Example 27. Partial Sequence Determination of Pollen-Specific

Proteins

Protein bands determined to be pollen-specific were

purified by electroblotting from the polyacrylamide gel onto

PVDF membrane (Matsudaira, P., _J. Biol. Chem._ 261:10035-10038

(1987)) or by reverse phase HPLC. N-terminal sequence of the

purified proteins was determined by automated Edman egradation

with an Applied Biosystems 470A gas-phase sequencer.

Phenylthiohydantoin (PTH) amino acids were identified using an

Applied Biosystems 120A PTH analyzer. To obtain internal

sequence, proteins were digested with endoproteinase Lys-C

(Boehringer Mannheim) in 0.1 M Tris-HCl, pH 8.5, for 24 hours

at room temperature using an enzyme:substrate ratio of 1:10.

Resulting peptides were isolated by HPLC using an Aquapore C-8

column eluted with a linear acetonitrile/isopropanol (1:1

ratio) gradient (0 to 60%) in 0.1% TFA. Sequence of isolated

Lys-C peptides was determined as above. The following

sequences were determined for the 13kD pollen-specific protein:

N-terminus: TTPLTFQVGKGSKPGHLILTPNVATI

LysC 61:   KPGHLILTPNVATISDVVIK

LysC 54:   SGGTRIADDVIPADFK

LysC 49:   EHGGDDFSFTLK

LysC 43:   EGPTGTWTLDTK

Example 28. Synthesis of Oligonucleotide Probes for

Pollen-Specific cDNAs

Regions of peptide sequence in the 13kD protein with low codon redundancy were selected, and suitable oligonucleotide probes for the gene encoding these regions were synthesized on an Applied Biosystems 380A synthesizer. The following oligonucleotides were synthesized:

```
Oligo #51      5'-AA ATC ATC ACC ACC ATG TTC-3'
                      G   G   G       G   C
                                  T
                                  C
```

```
Oligo # 58     5'-CC TTT ACC CAC TTG AAA-3'
                      C   G       C   G
                              T
                              C
```

where the columns of nucleotides represent bases that were incorporated randomly in equal proportions at the indicated position in the oligo. Oligo #51 encodes the amino acid sequence EHGGDDF found in peptide LysC 49, and Oligo #58 encodes the amino acid sequence FQVGKG found in peptide N-terminus. Use of these mixed oligonucleotides to screen a cDNA library for the pollen-specific gene will be described below.

Example 29. Construction of a maize pollen cDNA library

Total maize RNA from maize 211D shed pollen was

isolated as described in Glisen et al, <u>Biochemistry</u>
13:2633-2637 (1974).   Poly A+ mRNA was purified from total RNA
as described in Sambrook et al.   Using this mRNA, cDNA was
prepared using a cDNA synthesis kit purchased from Promega,
following protocols supplied with the kit.   The EcoRI linkers
were added to the cDNA and it was ligated into arms of the
cloning vector lambda Zap, purchased from Stratagene and using
the protocol supplied by the manufacturer.   The ligation
product was packaged in a lambda packaging extract also
purchased from Stratagene, and used to infect <u>E. coli</u> BB4
cells.

Example 30.   Isolation of pollen-specific cDNA clones

        . The maize pollen cDNA library was probed using the
synthetic oligonucleotides probes specific for the 13kD protein
gene, as described in Sambrook et al. Briefly, about 100,000
phage plaques of the pollen cDNA library were plated and lifted
to nitrocellulose filters.   The filters were probed using
oligonucleotides #51 and #58 which had been 32P end-labeled
using polynucleotide kinase.   The probes were hybridized to
the filters at low stringency (50 degrees C in 1M NaCl, 10%
dextran sulfate, 0.5% SDS), washed 30 minutes at room
temperature and then 30 minutes at 45 degrees C in 6X SSC, 0.1%
SDS, and exposed to X-ray film to identify positive clones.
Putative clones were purified through four rounds of plaque
hybridization.   Three classes of cDNA clones were isolated.
Type I contained EcoRI fragments of 0.2 kb and 1.8 kb.   Type II
contained EcoRI fragments of 0.6 kb, 0.5 kb and 1.0 kb, and

Type III contained an EcoRI fragment of 2.3 kb.

Example 31.   Characterization of Pollen-specific cDNA clones

          The EcoRI fragments of the Type II cDNA clone were

subcloned into the plasmid vector pBluescript SK+, purchased

from Stratagene. See Fig. 30. The 0.6 kb fragment in

pBluescript was named II-.6, the 0.5 kb fragment in pBluescript

was named II-.5 (later renamed pCIB3169) and the 1.0 kb

fragment in pBluescript was named II-1.0 (later renamed

pCIB3168). As will be described below, the 0.5 kb and 1.0 kb

fragments encode the maize pollen-specific CDPK gene. RNA from

anthers, pollen, leaf, root and silk was denatured with

glyoxal, electrophoresed on a 1% agarose gel, transferred to

nitrocellulose, and probed separately with the three EcoRI

fragments that had been labeled with 32P by random primer

extension as described in Sambrook et al, Molecular Cloning, A

Laboratory Manual, Cold Spring Harbor Laboratory Press: New

York (1989). The blots were exposed to X-ray film, and an mRNA

band of approximately 1.5 kb was identified with the 0.6 kb

fragment probe, while the 0.5 and 1.0 kb fragments hybridized

to an approximately 2.0 kb mRNA. In all cases hybridization

was only seen in the pollen RNA lane, with the exception that

the 0.6 kb fragment showed a slight signal in anther mRNA. The

conclusion from these data was that the original cDNA clone was

a fusion cDNA molecules derived from two different mRNAs. The

0.6 kb fragment was a partial cDNA of a 1.5 kb pollen-specific

mRNA, and this mRNA encodes the peptides LysC 49 and

N-terminus. The 1.0 and 0.5 kb fragments comprise a partial

cDNA of a 2.0 kb pollen-specific mRNA unrelated to the peptides and oligonucleotide probes used for probes. This conclusion was verified when the fragments were sequenced using the dideoxy chain termination method as described in Sambrook et al. The cDNA sequence is shown in Fig. 31.

Example 32. Determination of specificity of mRNA expression

To determine if the 2.0 kb RNA represented by cDNA clones pCIB3169 and pCIB3168 were present only in pollen, total RNA was isolated from maize 211D roots, leaves, pollen, anthers or silks. The RNAs were denatured with glyoxal, electrophoresed on a 1% agarose gel, transferred to nitrocellulose, and probed with 32P-labeled EcoRI insert from plasmid pCIB3168 or pCIB3169, all using standard techniques as described in Sambrook et al, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press: New York (1989). Exposure of this blot to photographic film demonstrates that the gene represented by these two clones is only transcriptionally active in the pollen (Fig. 32).

Identification of a Pollen-Specific Promoter

Example 33. Construction of a Maize Genomic DNA Library

Genomic DNA from maize line 211D young shoots was isolated as described in Shure et 1, Cell 35:225-233 (1983). The DNA was provided to Stratagene, where a genomic DNA library was constructed by cloning Sau3AI partially digested DNA into Stratagene's Lambda Dash cloning vector.

Example 34. Genomic DNA Blot Hybridization to Determine Gene Copy Number.

118

Genomic DNA from maize line 211D was digested with a

number of restriction enzymes, the individual digests

electrophoresed on an agarose gel, transferred to

nitrocellulose, and probed with 32P-labeled EcoRI insert from

plasmid pCIB3168 (1.0 kb fragment), pCIB3169 (0.5 kb fragment)

or clone II-.6 using standard techniques described in Sambrook

et al. More than 10 bands were detected by the II-.6 probe on

most digests, indicating that this cDNA is derived from a

large, multigene family. Probing with the 1.0 kb fragment

detected from 3 to 6 bands, and probing with the 0.5 kb

fragment detected only from 1 to 3 bands which were a subset of

those detected by the 1.0 kb fragment. Due to the smaller gene

family size detected by the 1.0 kb and 0.5 kb fragments, it was

decided to attempt to isolate the genomic clone corresponding

to them.

Example 35.   Isolation of a pollen-specific genomic clone

The Stratagene maize 211D genomic library was screened

by probing plaque lifts with 32P labeled inserts from plasmid

pCIB3168 (1.0 kb fragment) and pCIB3169 (0.5 kb fragment) using

standard procedures as described in the Stratagene manual

accompanying the library.   Using this strategy, Lambda clone

MG14 was isolated, and it hybridized to both probes. The 9.0

kb BamHI fragment of MG14, which also hybridized to both

probes, was subcloned into the BamHI site of pBluescript SK+ to

create plasmid pCIB379. 1800 bp of pCIB379, in the region

corresponding to the cDNA sequence, was sequenced as described

above. Comparison of the cDNA and genomic sequences showed

119

only 91% identity. pCIB379 insert represents a related pollen-specific gene.

A second maize 211D genomic library was constructed in the vector lambda GEM-11, purchased from Promega, using the procedures described in the Promega manual. Screening this un-amplified library as above yielded clone GEM11-1, which hybridized to both 0.5 and 1.0 kb probes. The 20kb HindIII fragment of GEM11-1, which also hybridized to both probes, was subcloned into the HindIII site of pBluescript SK+ to yield pCIB3166. The DNA sequence of 4.1kb of pCIB3166 was determined (Fig. 36) and after accounting for six introns in the genomic clone, was 100% identical to the cDNA sequence of pCIB3168 and pCIB3169. Comparison of the pCIB3166 sequence to the Genbank/EMBL database revealed that the 5' portion, through the 3 exon, was 34.6% identical to rat calmodulin-dependent protein kinase II at the amino acid level (Fig. 33), while the fourth through seventh exons were 39.4% identical to rat calmodulin. See Fig. 34. No other pollen-specific kinase has been described, and at the time this a protein combining kinase and calmodulin domains was unknown. Subsequently, Harper et al., Science 252:951-954 (1991) have disclosed the cDNA sequence of a similar protein from soybean, although this gene is not pollen-specific in expression. Comparison of the soybean calcium-Dependent Protein Kinase (CDPK) and the maize pollen CDPK reveals 38% identity at the amino acid level. See Fig. 35.

Example 36. Identification of the Promoter's Transcriptional

Start Site by Primer Extension

Oligonucleotide PE51, with the following sequence was synthesized as a primer.

5'-TGGCCCATGGCTGCGGCGGGGAACGAGTGCGGC-3'

Primer extension analysis was carried out on polyA+ pollen mRNA as described in Metraux et al., PNAS USA 86:896-890 (1989). The transcription initiation site was determined to be between bases 1415 and 1425 on the partial sequence of pCIB3166 shown in Fig. 36.

Testing Promoter Function in Transgenic Plants
Example 37. Construction of promoter vectors for plant transformation

To demonstrate that the pollen CDPK promoter can drive expression of a linked gene in transgenic plants, a gene fusion of the pollen CDPK promoter to the Beta-glucuronidase gene of E. coli was constructed as follows. The 10kb BamHI fragment from lambda GEM11-1 containing the first exon and part of the first intron of the pollen CDPK gene plus 9 kb upstream of the gene was subcloned into the BamHI site of pBluescript SK+ to create plasmid pCIB3167. The 2.3 kb BamHI-HindIII fragment from pCIB3167 was subcloned into the BamHI and HindIII sites of pBluescript SK+ to create plasmid pSK105. The pSK105 was digested with AvaI and HindIII, and the1.75 kb HindIII-AvaI fragment was isolated on an agarose gel. A PCR reaction was run under standard conditions as described in Sambrook et al. using intact pSK105 as a template and the following primers:

#42: 5'-AGCGGTCGACCTGCAGGCATGCGATCTGCACCTCCCGCCG-3'

121

#43:   5'-ATGGGCAAGGAGCTCGGG-3

The PCR reaction products were digested with AvaI and SalI and the resulting fragment isolated on an agarose gel. pBluescript SK+ was digested with HindIII and SalI. The 1.75 kb HindIII-AvaI fragment, PCR derived AvaI-SalI fragment, and pBluescript vector with HindIII and SalI ends were ligated in a three way ligation to create plasmid pSK110.

A fusion of the promoter fragment in pSK110 to the Beta-glucuronidase (GUS) gene was created by digesting pSK110 with HindIII and SalI, isolating the 1.9kb fragment on an agarose gel and ligating it into HindIII and SalI sites of pCIB3054, to create plasmid pKL2, a plasmid derived from pUC19 containing the GUS gene followed by plant intron from the maize PEPC gene and a polyA signal from cauliflower mosaic virus. This promoter fusion was inactive in plants, probably due to the presence of out of frame ATG codons in the leader sequence preceding the GUS gene ATG.

A function fusion of the promoter was created by digesting pKL2 with XbaI and SalI to remove the previous fusion junction. A new fusion junction was produced in a PCR reaction using pSK105 as a template and the following primers:

#SK50: 5'-CCCTTCAAAATCTAGAAACCT-3'

#SK49: 5'-TAATGTCGACGAACGGCGAGAGATGGA-3'

The PCR product was digested with XbaI and SalI and purified on an agarose gel. The purified fragment was ligated into the XbaI and SalI sites of pKL2 to created plasmid pCIB3171. This plasmid contains a functional fusion of pollen

122

CDPK promoter and GUS which directs expression the GUS gene
exclusively in pollen.

To create a vector containing the pollen CDPK
promoter-GUS fusion suitable for use in _Agrobacterium_
tumefaciens-mediated plant transformation, the fusion gene was
isolated from pCIB3171 by digestion with HindIII and SalI. The
resulting fragment was ligated into the HindIII and SalI sites
of pBI101 (purchased from Clontech) to create plasmid pCIB3175.

Example 38. Production of Transgenic Plants

pCIB3175 was transformed into Agrobacterium tumefaciens
containing the helper plasmid pCIB542, and the resulting
culture used to transform leaf disks from tobacco shoot tip
cultures as described by Horsch et al., _Science_ 227:1229-1231
(1985) except that nurse cultures were omitted and selection
was on 100 mg/l kanamycin. Transgenic plants were regenerated
and verified for presence of the transgene by PCR.

Example 39. GUS Gene Expression Analysis

Pollen from primary transformants and their progeny
were analyzed histochemically for expression of the GUS gene as
described by Guerrero et al., _Mol. Gen. Genet._ 224:161-168
(1990). The percentage of pollen grains expressing the GUS
gene, as demonstrated by blue staining in the X-gluc buffer, is
shown in the table below.

Plant Number	% Blue Pollen
PP1-51	28%
PP1-54	54%
PP1-55	none
PP1-61	very few
PP1-63	51%

123

PP1-67   15%
PP1-80   10%
PP1-83   12%

Primary transformants in which a single pollen CDPK promoter-GUS gene was integrated would produce a maximum 50% GUS positive pollen due to segregation of the single gene.

Flouometric GUS assays were done on pollen, stem, root, leaf and pistil tissue of selected plants to demonstrate the specificity of pollen CDPK promoter expression. Assays were performed as described in Jefferson, Plant Mol. Biol. 14:995-1006 (1990), and GUS activity values are expressed as nmoles MU/ug protein/minute.

Plant number	Tissue	GUS Activity	Untransformed Plant GUS Activity	Net GUS Activity
PP1-51	stem	0.01	0.02	0
	leaf	0	0	0
	root	0.15	0.10	0.05
	pistil	0.02	0.01	0.01
	pollen	0.24	0.02	0.22
PP1-54	stem	0.01	0.02	0
	leaf	0	0	0
	root	0.13	0.1	0.03
	pistil	0.01	0.01	0
	pollen	0.60	0.02	0.58
PP1-63	stem	0.01	0.02	0
	leaf	0	0	0
	root	0.07	0.1	0
	pistil	0.01	0.01	0
	pollen	0.57	0.02	0.55

Examples 40-50 are directed primarily to the preparation of chimeric constructs, i.e. recombinant DNA molecules, containing constitutive, tissue-preferred, or tissue-specific promoters operably linked to an instant B.t. gene, insertion of same into

vectors, production of transgenic platns containing the vectors, and analysis of expression levels of <u>B.t.</u> proteins of the transgenic plants.

EXAMPLE 40: CONSTRUCTION OF MAIZE OPTIMIZED BT
<u>TRANSFORMATION VECTORS</u>

To demonstrate the effectiveness of the synthetic Bt cryIA(b) gene in maize, the PepC and pith specific promoters are fused to the synthetic Bt cryIA(b) gene using PCR. Oligomers designed for the PCR fusions were:

(PEPC)

KE99A28 = 5'-TGCGGTTACC GCCGATCACATG-3'

KE97A28 = 5'-GCGGTACCGC GTCGACGCGG ATCCCGCGGC GGGAAGCTAAG-3'

(PITH)

KE100A28 = 5'-GTCGTCGACC GCAACA-3'

KE98A28 = 5'-GCGGTACCGC GTTAACGCGG ATCCTGTCCG ACACCGGAC-3'

KE104A28 = 5'-GATGTCGTCG ACCGCAACAC-3'

KE103A28 = 5'-GCGGTACCGC GGATCCTGTC CGACACCGGA CGGCT-3'

PCR primers are designed to replace the Nco I sites in the 5' untranslated leader region of each of these tissue specific genes (containing ATG translational start sites) with Bam HI sites to facilitate cloning of the synthetic cryIA(b) gene into this Bam HI site. Subsequent construction of vectors

containing the tissue specific promoters fused to the synthetic cryIA(b) gene and also containing the 35S:PAT:35S marker gene involves several intermediate constructs.

1. pCIB4406 (35S:synthetic-cryIA(b):pepC ivs#9:35S)

pCIB4406 contains the 2 Kb Bam HI\Cla I synthetic cryIA(b) gene fused with the CaMV 35S promoter (Rothstein et al., Gene 53:153-161 (1987)). The gene also contains intron #6 derived from the maize PEP carboxylase gene (ivs#9) in the 3' untranslated region of the gene, which uses the CaMV 3' end. (PNAS USA, 83:2884-2888 (1986), Hudspeth et al., Plant Molecular Biology, 12: 579-589 (1989)). pCIB4406 is ligated and transformed into the "SURE" strain of E. coli cells (Stratagene, La Jolla, CA) as described above. One mutation is found in pCIB4406's cryIA(b) gene at amino acid #436 which resulted in the desired Phe being changed to a Leu. pCIB4406 is fully active against European corn borer when tested in insect bioassays and produces a CryIA(b) protein of the expected size as determined by western blot analysis.

2. pCIB4407 (35S:synthetic-cryIA(b):pepC ivs#9:35S + 35S:PAT:35S)

pCIB4407 is made from an approximately 4 Kb Hind III\Eco RI fragment containing the 35S:PAT:35S gene, and the 3.1 Kb\Hind III\Eco RI 35S:synthetic-cryIA(b):35S gene from pCIB4406. pCIB4407 is ligated and transformed into "SURE", DH5alpha, and HB101 strains of E. coli using standard procedures (Sambrook et al.). The synthetic cryIA(b) gene has

126

the same properties as its precursor pCIB4406.

3. pCIB4416 (35S:synthetic-cryIA(b):pepC ivs#9:35S +

      35S:PAT:35S + 35S:Adh intron:GUS:35S.)

    pCIB4407 is cut with Eco RI and treated with calf
intestinal alkaline phosphatase (CIP) under standard conditions
(Sambrook et al.) to produce an about 7.2 Kb fragment that is
ligated with a 3.4 Kb Eco RI 35S:Adh\GUS:35S fragment to
produce pCIB4416. Ligations and transformations into "SURE"
cells is as described above. The synthetic cryIA(b) gene in
pCIB4416 has the same properties as the gene in pCIB4406.

4. pCIB4418 (35S:synthetic-cryIA(b):pepC ivs#9:35S)

    pCIB4406 is digested with Apa I and Bam HI and treated
with CIP. pCIB4406 is digested with Bam HI and Nsp I.
pBS123#13 is digested with Nsp I and Apa I. A three-way
ligation is made consisting of a 4.3 Kb Apa I\Bam HI fragment
from pCIB4406, a 1.3 Kb Bam HI\Nsp I fragment from pCIB4406,
and a 170 bp Nsp I\Apa I fragment from pBS123#13 to form
pCIB4418. The host E. coli strain for pCIB4418 is HB101.

5. pCIB4419 (35S:synthetic-cryIA(b):pepC ivs#9:35S +

      35S:PAT:35S + 35S:Adh intron:GUS:35S.)

    pCIB4416 and pCIB4418 are digested with Bst E II and
Eco NI and fragments of pCIB4416 are treated with CIP. A 9.1
Kb fragment from pCIB4416 ligated to a 1.4 Kb fragment from
pCIB4418 to form pCIB4419. pCIB4419 transformed in HB101
competent E. coli cells demonstrates full activity in insect
bioassays against European corn borer.

6. pCIB4420 (Pith:synthetic-cryIA(b):PEPC ivs#9:35S +

35S:PAT:35S)

Intermediate constructs in making pCIB4420 are pBTin1,

pBtin2, p4420A and pBtin3. pBtin1 (pith promoter:second half

of the synthetic Bt gene + 35S:PAT:35S) is made by ligating the

2.1 Kb Xba I\Nco I pith promoter fragment from plasmid

pith(3-1) with a 5.2 Kb Xba I\Nco I fragment from pCIB4407.

pBtin2 is an intermediate construct containing the pith

promoter modified with a 210 bp PCR fragment made using primers

KE100A28 and KE98A28 listed above. The PCR reaction mix

contains approximately 100 ng of a 2.1 Kb Bam HI\Nco I pith

promoter fragment with 100 pmol of each oligomer, 200 nM of

each dNTP, 1 X buffer (Cetus) and 2.5 units of thermal stable

polymerase. Since the Tm is relatively low (between 40° and

50°C), PCR reactions are run with the following parameters:

denaturation cycle: 94°C for 1 minute

annealing cycle : 37°C for 1 minute

extension cycle : 72°C for 45 seconds (+ 3 seconds per

cycle)

number of cycles: 25

PCR reactions are treated with proteinase K as

described above prior to cutting with Sal I\Kpn I followed by

phenol\chloroform extraction and ethanol precipitation as

described above. The 210 bp fragment is purified on a 2 %

Nusieve gel and extracted from the gel using Millipore's filter

units. The 210 bp Sal I\Kpn I fragment is ligated to the 4.9

Kb Sal I\Kpn I fragment from pith(3-1) to make pBtin2. p4420A
(pith:synthetic-Bt:Pep intron:35S + 35S:PAT:35S) is made with a
three-way ligation consisting of a 700 bp Nsi I\Bam HI fragment
from pBtin2, a 1.8 Kb Bam HI\Bst E II fragment from pCIB4418,
and a 5.9 Kb Bst E II\Nsi I fragment from pBtin1. After p4420A
is made three mutations are discovered in pBtin2. A second PCR
fragment is made to modify the Nco I site in the pith leader
using primers KE104A28 and KE103A28 with Tm values around 65°C.
The PCR reaction mix is identical to that listed above with the
addition of glycerol to 20% to reduce mutations in G+C rich
areas (Henry et al., <u>Plant Molecular Biology Reporter</u>
9(2):139-144, 1991). PCR parameters are as follows:
File I: 94°C : 3 minutes , 1 cycle
File II: 60°C : 1 minute

           94°C : 1 minute

           25 cycles
File III: 72°C : 5 minutes, 1 cycle

     PCR reactions are treated as above and cut with
restriction endonucleases Sal I and Kpn I. The 210 bp Sal
I\Kpn I PCR (glycerol in the reaction) fragment is ligated to
the 4.9 Kb Sal I\Kpn I fragment from plasmid pith(3-1) to make
pBtin3. Sequence data on pBtin3-G#1 shows this PCR generated
fragment to be correct.
     pBtin3-G#1 is used to make pCIB4420 (also called p4420B
"G#6"). pCIB4420 is constructed with a three-way ligation
using the 700 bp Nsi I\Bam HI fragment from pBtin3-G#1, a 1.8

Kb Bam HI\Bst E II fragment from pCIB4418, and a 5.9 Kb Bst E

II\Nsi I fragment from pBtin1. pCIB4420 is used in mesophyll

protoplast experiments and demonstrates full activity of the

synthetic cryIA(b) gene against European corn borer.

7. pCIB4413 (PEPC:synthetic-Bt (Phe mutation):PEPC intron:35S.)

A fusion fragment is generated by PCR using primers

KE99A28 and KE97A28 with a 2.3 KB Hind III\Sal I template from

pGUS4.5. The PCR mix contains the same concentration of

primers, template, dNTPs, salts, and thermal stable polymerase

as described above. PCR reaction parameters are:

denaturation cycle : 94°C for 1 minute

annealing cycle : 55°C for 1 minute

extension cycle : 72°C for 45 seconds (+ 3 seconds per

cycle)

number of cycles: 30

After completion, PCR reactions are treated with

proteinase K followed by phenol\chloroform extraction and

ethanol precipitation as described above prior to cutting with

restriction endonucleases Bam HI and Bst E II.

pCIB4413 is made with a three-way ligation using the

210 bp Bam HI\Bst E II PCR fragment, a 4.7 Kb Bam HI\Hind III

fragment from pCIB4406, and a 2.2 Kb Hind III\Bst E II fragment

from pGUS4.5.

8. pCIB4421 (PEPC:synthetic-cryIA(b):PEPC intron:35S.)

pCIB4421 is made to replace the synthetic cryIA(b) gene

containing the Phe mutation in pCIB4413 with the synthetic

cryIA(b) gene from pCIB4419. pCIB4421 is made by ligating a

5.2 Kb Bam HI\Sac I fragment from pCIB4413 with a 1.9 Kb Bam

HI\Sac I fragment from pCIB4419.

9. pCIB4423 (PEPC:synthetic-cryIA(b):PepC intron:35S +

35S:PAT:35S)

   The 2.4 Kb Bam HI\Hind III PEPC promoter fragment from

pCIB4421 is ligated to the 6.2 Kb Bam HI\Hind III fragment in

pCIB4420 to make pCIB4423. The Hind III site is deleted by

exonucleases in the cloning of pCIB4423. pCIB4423 contains the

synthetic cryIA(b) gene under the control of the PEPC promoter,

and the PAT gene under the control of the 35S promoter.

10. Synthetic cryIA(b) gene in Agrobacterium strains:

   Agrobacterium strains made with the synthetic cryIA(b)

gene allow transfer of this gene in a range of dicotyledenous

plants. Agrobacterium vector pCIB4417 contains the 3.3 Kb Hind

III\Eco RI 35S:synthetic-CryIA(b):PepC:ivs#9:35S fragment from

pCIB4406 (Phe mutation) ligated to the 14 Kb Hind III\Eco RI

fragment from pBI101 (Clontech). Using electroporation,

pCIB4417 is transferred into the A. tumefaciens strain LBA4404

(Diethard et al., Nucleic Acids Research, Vol17:#16:6747,

1989.).

   200 ng of pCIB4417 and 40 ul of thawed on ice LBA4404

competent cell are electroporated in a pre-cooled 0.2 cm

electroporation cuvette (Bio-Rad Laboratories Ltd.). Using

Gene Pulser-TM with the Pulse Controller unit (Bio-Rad), an

electric pulse is applied immediately with the voltage set at

2.5 kV, and the capacity set at 25 uF. After the pulse, cells
are immediately transferred to 1 ml of YEB medium and shaken at
27 C for 3 hours before plating 10 ul on ABmin:Km50 plates.
After incubating at 28 C for approximately 60 hours colonies
are selected for miniscreen preparation to do restriction
enzyme analysis. The final Agrobacterium strain is called
pCIB4417:LBA4404.

EXAMPLE 41: ELISA ANALYSIS OF TRANSFORMED MAIZE PROTOPLASTS

The presence of the cryIA(b) toxin protein is detected
by utilizing enzyme-linked immunosorbent assay (ELISA). ELISAS
are very sensitive, specific assays for antigenic material.
ELISA assays are useful to determine the expression of
polypeptide gene products. Antiserum for these assays is
produced in response to immunizing rabbits with
gradient-purified Bt crystals [Ang et al., Applied Environ.
Microbiol., 36:625-626 (1978)] solubilized with sodium dodecyl
sulfate. ELISA analysis of extracts from transiently
transformed maize cells is carried out using standard
procedures (see for example Harlow, E., and Lane, D. in
"Antibodies: A Laboratory Manual", Cold Spring Harbor
Laboratory Press, 1988). ELISA techniques are further
described in Clark et al., Methods in Enzymology, 118:742-766
(1986); and Bradford, Anal. Biochem., 72:248 (1976). Thus,
these procedures are well-known to those skilled in the art.
The disclosure of these references is hereby incorporated
herein by reference.

ELISA assays are performed to detect the production of

132

CryIA(b) protein in maize protoplasts. Protein produced is
reported below as ng of Bt per mg total protein (ng Bt/mg).
Each construct was tested twice.

pCIB3069   No detectable Bt (both tests)

pCIB4407   21,900 ng Bt/mg total protein,

           21,000 ng Bt/mg total protein

The transformed maize cells produce high levels, on the
order of approximately 20,000 ng of Bt CryIA(b) protein per mg
total soluble protein, of the Bt IP when transformed with the
maize optimized Bt gene. The level of detection of these ELISA
based assays is about 1 to 5 ng CryIA(b) protein per mg
protein. Therefore, the maize optimized Bt gene produces as
much as approximately a 20,000 fold increase in expression of
this protein in maize cells.

EXAMPLE 42: ASSAY OF EXTRACT FROM TRANSFORMED PROTOPLASTS FOR INSECTICIDAL ACTIVITY AGAINST EUROPEAN CORN BORER

Western blot analysis is also performed using extracts
obtained from maize cells which had been transiently
transformed with DNA to express the maize optimized gene. When
examined by western blots, this protein appears identical with
the protein produced in E. coli. In contrast, as demonstrated
in Example 6 above, no detectable Bt cryIA(b) insecticidal
protein is produced by maize cells transformed with comparable
vectors attempting to express the native Bt derived coding
region.

Qualitative insect toxicity testing can be carried out

133

using harvested protoplasts. Suspensions are prepared for each replicate tested in all bioassays. A replicate is considered positive if it causes significantly higher mortality than the controls. For example, replicates are tested for their activity against insects in the order Lepidoptera by using the European corn borer, Ostrinia nubilalis. One-hundred μl of a protoplast suspension in 0.1% Triton X-100 is pipetted onto the surface of artificial Black cutworm diet, (Bioserv, Inc., Frenchtown, NJ; F9240) in 50 mm X 10 mm snap-cap petri dishes. After air drying 10 neonatal larvae are added to each plate. Mortality is recorded after about 4 days. When this protein is fed to European corn borers, it produces 100% mortality.

EXAMPLE 43: EXPRESSION OF SYNTHETIC BT IN MAIZE MESOPHYLL PROTOPLASTS

The general procedure for the isolation of corn mesophyll protoplasts is adapted from Sheen et al., The Plant Cell, 2:1027-1038 (1990). The protoplast transformation system used in Sheen et al. is modified by using PEG mediated transformation, rather than electroporation. That procedure, as well as changes made in the isolation procedure, is described below.

Maize Mesophyll Protoplast Isolation/Transformation

1. Sterilize and germinate corn seeds for leaf material. Seedlings are grown in the light at 25C.

2. Surface sterilize leaf pieces of 10-12 day old seedlings with 5% Clorox for 5 minutes followed by several

washes with sterile distilled water.

3. Aliquot enzyme solution (see recipe below); 25 ml/dish (100x25 mm petri dish).

4. Remove any excess water from leaves and place 6-8 2 inch pieces in each dish of enzyme. 14 plates are usually set up with the leaf material from about 100 seedlings.

5. Cut leaves in longitudinal strips as thin as possible (2-5 mm).

6. Shake slowly at 25C for 6.5 to 7 hours. Cover plates so that incubation takes place in the dark.

7. Before filtering protoplasts, wash 100 um sieves with 10 ml 0.6 M mannitol. Pipet protoplasts slowly through sieves. Wash plates with 0.6 M mannitol to gather any protoplasts left in the dishes.

8. Pipet filtered liquid carefully into 50 ml sterile tubes. Add equal volumes of 0.6 M mannitol to dilute.

9. Spin for 10 minutes at 1000 rpm/500 g in table-top centrifuge (Beckman Model TJ-6).

10. Remove enzyme solution and discard. Resuspend pellets carefully in 5 ml mannitol. Pool several pellets. Bring volume to 50 ml with 0.6 M mannitol and spin.

11. Resuspend to a known volume (50 ml) and count.

12. After counting and pelleting, resuspend protoplasts at 2 million/ml in resuspending buffer (recipe below). Allow ppts to incubate in the resuspending buffer for at least 30 min before transformation.

Transformation:

135

1. Aliquot plasmids to tubes (Fisherbrand polystyrene 17 x 100 mm Snap Cap culture tubes); at least three replicates per treatment; use equimolar amounts of plasmids so that equal gene copy numbers are compared.

2. Add 0.5 ml protoplasts and 0.5 ml 40% PEG made with 0.6 M mannitol.

3. Shake gently to mix and incubate at 25C for 30 min.

4. Add protoplast culture media at 5 min intervals: 1,2,5 ml

5. Spin for 10 min at 1000 rpm/500 g.

6. Remove liquid from pellet and resuspend in 1 ml culture media (BMV media)

7. Incubate overnight at 25C in the dark.

Recipes:

Enzyme Solution

0.6 M mannitol

10 mM MES, pH 5.7

1 mM $CaCL_2$

1 mM $MgCl_2$

0.1% BSA

filter-sterilize

To this solution, add the following enzymes:

1% Cellulase RS, and 0.1% Macerozyme R10

Wash Buffer: 0.6 M mannitol, filter-sterilize

Resuspending Buffer: 0.6 M mannitol, 20 mM KCl,

filter-sterilize

Culture Media: BMV media recipe from:

Okuno et al., Phytopathology 67:610-615 (1977).

0.6 M mannitol

4 mM MES, pH 5.7

0.2 mM KH_2PO_4

1 mM KNO_3

1 mM $MgSO_4$

10 mM $CaCl_2$

1X K3 micronutrients

filter-sterilize

ELISA analysis of transformed protoplasts is done one
day after transformation. ELISA's are done as previously
described. The following three experiments are done with maize
inbred line 211D. Of course, other lines of maize may be used.
50 ug of plasmid pCIB4419 and equimolar amounts of other
plasmids are used. Total soluble protein is determined using
the BioRad protein assay. (Bradford, Anal.Biochem, 72:248
(1976).

Transformation Experiment:

Constructs tested:

1. pCIB4419 (Construct contains synthetic Bt under control of
CaMV 35S promoter and 35S/PAT and 35S/GUS marker genes)

2. pCIB4420 (Construct contains synthetic Bt under control of

Pith promoter and PAT marker gene)

3. pCIB4421 (Construct contains synthetic Bt under control of
PEPC promoter)

4. pCIB4423 (Construct contains synthetic Bt under control of
PEPC promoter and PAT marker gene)

(PEPC:synthetic-cryIA(b):PepC intron:35S + 35S:PAT:35S)

    In the following experiments, 10 or 11 day old 211D
seedlings are analyzed for production of the Bt CryIA(b)
protein in the Biorad protein assay:

Experiment 1 (11 day seedlings):

pCIB4419   15,000 ± 3,000 ng Bt/mg protein

pCIB4420   280 ± 65 ng Bt/mg protein

pCIB4421   9,000 ± 800 ng Bt/mg protein

Experiment 2 (10 day seedlings):

pCIB4419   5,000 ± 270 ng Bt/mg protein

pCIB4420   80 ± 14 ng Bt/mg protein

pCIB4421   1,600 ± 220 ng Bt/mg protein

Experiment 3 (11 day seedlings):

pCIB4419   21,500 ± 1,800 ng Bt/mg protein

pCIB4420   260 ± 50 ng Bt/mg protein

pCIB4421   11,900 ± 4,000 ng Bt/mg protein

pCIB4423   7,200 ± 3,400 ng Bt/mg protein

The above experiments confirm that both the CaMV 35S and PEPC promoters express the synthetic Bt CryIA(b) protein at very high levels. The pith promoter, while less efficient, is also effective for the expression of synthetic CryIA(b) protein.

EXAMPLE 44: STABLE EXPRESSION OF SYNTHETIC BT IN LETTUCE

The synthetic Bt gene in the Agrobacterium vector pCIB4417 is transformed into Lactuca sativa cv. Redprize (lettuce). The transformation procedure used is described in Enomoto et al., Plant Cell Reports, 9:6-9 (1990).
Transformation procedure:

Lettuce seeds are suface sterilized in 5% Clorox for 5 minutes followed by several washes in sterile distilled water. Surface-sterilized seeds are plated on half strength MS media (Murashige and Skoog, Physiol. Plant. 15:473-497 (1962)). Cotyledons of 6-day-old Redprize seedlings, grown under illumination of 3,000 lx 16 hr at 25C, are used as the explants for Agrobacterium infection. The base and tip of each cotyledon are removed with a scalpel. The explants are soaked for 10 minutes in the bacterial solution which have been cultured for 48 hours in AB minimal media with the apropriate antibiotics at 28C. After blotting excess bacterial solution

on sterile filter paper, the explants are plated on MS media (0.1 mg/l BA and 0.1 mg/l NAA) for 2 days. Explants are then transferred to selective media containing 500 mg/l carbenicillin and 50 mg/l kanamycin. The explants are subcultured to fresh media weekly. The growth chamber conditions are 16 hour 2,000 lx light at 25C. After approximately 4 weeks, an ELISA is done on healthy looking callus from each of four plates being subcultured. The ELISA procedure is the same as described above for protoplasts; soluble protein is again determined by the Biorad assay described above.

Results:

pCIB3021 (kan control)	0
pCIB4417 (plate 1)	0
pCIB4417 (plate 2)	505 ng Bt/mg protein
pCIB4417 (plate 3)	45 ng Bt/mg protein
pCIB4417 (plate 4)	1,200 ng Bt/mg protein

This example demonstrates that dicot plants can also show increased expression of the optimized insecticidal gene.

Example 45. Construction of pCIB4429.

pCIB4429 contains a preferred maize pollen-specific promoter fused with the maize optimized cryIA(b) gene. The pollen-specific maize promoter used in this construct was obtained from the plasmid pKL2, described in Example 37. The maize optimized cryIA(b) gene was obtained from plasmid pCIB4418, also described in Example 37.

pKL2 is a plasmid that contains a preferred maize

140

pollen-specific promoter fused with the E. coli
beta-glucuronidase gene. It was constructed from plasmids
pSK110 and pCIB3054. pSK110 contains the pollen specific maize
promoter. pCIB3054, a pUC19 derivative, contains the E. coli
beta-glucuronidase (GUS) gene fused with the cauliflower mosaic
virus (CaMV) 35S promoter. It's construction is described
elsewhere in this application. This promoter can be removed
from this plasmid by cutting with SalI/HindIII to yield a
fragment containing the GUS gene, a bacterial ampicillin
resistance gene and a ColEI origin of replication. A second
fragment contains the CaMV 35S promoter.

pCIB3054 was cut with the restriction enzymes SalI and
HindIII, using standard conditions, for 2 hours at room
temperature. The reaction was then extracted with
phenol/chloroform using standard conditions and the DNA
recovered by ethanol precipitation using standard conditions.
The recovered DNA was resuspended in buffer appropriate for
reaction with calf intestinal alkaline phosphatase (CIP) and
reacted with 2.5 units of CIP at 37°C overnight. After the CIP
reaction, the DNA was purified on an agarose gel using standard
conditions described elsewhere in this application. pSK110 was
cut with SalI/HindIII under standard conditions for 2 hours at
room temperature and the DNA subsequently purified on an
agarose gel using standard conditions. The recovered DNA
fragments were ligated using standard conditions for two hours
at room temperature and subsequently transformed into competent
E. coli strain HB101 cells using standard conditions.

Transformants were selected on L-agar containing 100 µg

ampicillin/ml. Transformants were characterized for the desired

plasmid construct using standard plasmid mini-screen

procedures. The correct construct was named pKL2.

To make pCIB4429, a three way ligation was performed

using standard conditions known to those in the art. The three

fragments ligated were:

1) a HindIII/BamHI fragment from pCIB4418, of about 4.7

kb in size, containing the cryIA(b) gene, the bacterial

ampicillin resistance gene, and the ColEI origin of replication

2) a HindIII/XbaI fragment from pKL2 of about 1.3 kb in

size and containing the pollen specific promoter from maize

3) a PCR generated fragment derived from the pollen

promoter with a BamHI site introduced downstream from the start

of transcription. This fragment is approximately 120 bp and

has ends cut with the restriction enzymes XbaI/BamHI.

The PCR fragment was generated using a 100 µl reaction

volume and standard conditions described above. The primers

used were:

SK50: 5'-CCC TTC AAA ATC TAG AAA CCT-3'

KE127: 5'-GCG GAT CCG GCT GCG GCG GGG AAC GA-3'

The above primers were mixed in a PCR reaction with

plasmid pSK105, a plasmid that contains the pollen specific

promoter from maize.

After the PCR reaction was complete, 10 µl of the

reaction was run on an agarose gel, using standard condition,

to make sure the reaction produced the expected size product.

The remaining 90 µl was treated with proteinase K at a final concentration of 50 µg/ml for 30 min. at 37°C. The reaction was then heated at 65°C for 10 min., then phenol/chloroform extracted using standard procedures. The DNA was recovered from the supernatant by precipitating with two volumes of ethanol using standard conditions. After precipitation, the DNA was recovered by centrifuging in a microfuge. The pellet was rinsed one time with 70% ethanol (as is standard in the art), briefly dried to remove all ethanol, and the pellet resuspended in 17 µl TE buffer. 2 µl of 10X restriction enzyme buffer was added as were 0.5 µl BamHI and 0.5 µl XbaI. The DNA was digested for 1 hour at 37°C to produce a DNA fragment cut with XbaI/BamHI. After digestion with the restriction enzymes, this fragment was purified on an agarose gel composed of 2% NuSieve (FMC)/1% agarose gel. Millipore filter units were used to elute the DNA from the agarose using the manufacturer's specifications. After elution, the DNA was used in the three-way ligation described above.

After ligation, the DNA was transformed into competent E. coli strain HB101 cells using standard techniques. Transformants were selected on L-agar plates containing ampicillin at 100 µg/ml. Colonies that grew under selective conditions were characterized for plasmid inserts using techniques standard in the art.

EXAMPLE 46. Construction of pCIB4431, a vector for tissue specific expression of the synthetic cryIA(b) gene in plants.

pCIB4431 is a vector designed to transform maize. It

143

contains two chimeric Bt endotoxin genes expressible in maize. These genes are the PEP carboxylase promoter/synthetic-cryIA(b) and a pollen promoter/synthetic-cryIA(b). The PEP carboxylase/cryIA(b) gene in this vector is derived from pCIB4421 described above. The pollen promoter is also described above. Fig. 20 is a map of plasmid pCIB4431. pCIB4431 was constructed via a three part ligation using the about 3.5 Kb Kpn I/Hind III fragment (containing pollen/synthetic-cryIA(b) from pCIB4429, the about 4.5 Kb Hind III/Eco RI (PEPC/synthetic-cryIA(b) and the about 2.6 Kb Kpn I/Eco RI fragment from the vector Bluescript.

Other vectors including the pollen promoter/synthetic CryIA(b) chimeric gene include pCIB4428 and pCIB4430. See Figs. 21 and 22. pCIB4430 also contains the PEPC/synthetic-Bt gene described above.

EXAMPLE 47. Production of transgenic maize plants containing the synthetic maize optimized CryIA(b) gene

The example below utilizes Biolistics to introduce DNA coated particles into maize cells, from which transformed plants are generated.

Experiment KC-65

Production of transgenic maize plants expressing the synthetic cryIA(b) gene using a tissue-specific promoter.

Tissue

Immature maize embryos, approximately 1.5-2.5 mm in length, were excised from an ear of genotype 6N615 14-15 days after pollination. The mother plant was grown in the greenhouse.

144

Before excision, the ear was surface sterilized with 20% Clorox

for 20 minutes and rinse 3 times with sterile water.

Individual embryos were plated scutellum side in a 2 cm square

area, 36 embryos to a plate, on the callus initiation medium,

2DG4 + 5 chloramben medium (N6 major salts, B5 minor salts, MS

iron, 2% sucrose, with 5 mg/l chloramben, 20 mg/l glucose, and

10 ml G4 additions (Table 1) added after autoclaving.

TABLE 1 - G4 Additions

Ingredient	per liter medium	
Casein hydrolysate	0.5	gm
Proline	1.38	gm
Nicotinic acid	.2	mg
Pyridoxine-HCl	.2	mg
Thiamine-HCl	.5	mg
Choline-HCl	.1	mg
Riboflavin	.05	mg
Biotin	.1	mg
Folic acid	.05	mg
Ca pantothenate	.1	mg
p-aminobenzoic acid	.05	mg
B12	.136	µg

Bombardment

Tissue was bombarded using the PDS-1000He Biolistics

device. The tissue was placed on the shelf 8 cm below the

stopping screen shelf. The tissue was shot one time with the

DNA/gold microcarrier solution, 10 µl dried onto the

macrocarrier. The stopping screen used was hand punched at

ABRU using 10x10 stainless steel mesh screen. Rupture discs of

1550 psi value were used. After bombardment, the embryos were

cultured in the dark at 25° C.

Preparation of DNA for delivery

145

The microcarrier was prepared essentially according to the instructions supplied with the Biolistic device. While vortexing 50 µl 1.0 µ gold microcarrier, added 5 µl pCIB4431 (1.23 µg/µl) [#898] + 2 µl pCIB3064 0.895 µg/µl) [#456] followed by 50 µl 2.5 M $CaCl_2$, then 20 µl 0.1 M spermidine (free base, TC grade). The resulting mixture was vortexed 3 minutes and microfuged for 10 sec. The supernatant was removed and the icrocarriers washed 2 times with 250 µl of 100% EtOH (HPLC grade) by vortexing briefly, centrifuging and removing the supernatant. The microcarriers are resuspended in 65 µl 100% EtOH.

Callus formation

Embryos were transferred to callus initiation medium with 3 mg/l PPT 1 day after bombardment. Embryos were scored for callus initiation at 2 and 3 weeks after bombardment. Any responses were transferred to callus maintenance medium, 2DG4 + 0.5 2,4-D medium with 3 mg/L PPT. Callus maintenance medium is N6 major salts, B5 minor salts, MS iron, 2% sucrose, with 0.5 mg/l 2,4-D, 20 mg/l glucose, and 10 ml G4 additions added after autoclaving. Embryogenic callus was subcultured every 2 weeks to fresh maintenance medium containing 3 mg/L PPT. All callus was incubated in the dark at 25°C.

The Type I callus formation response was 15%. Every embryo which produced callus was cultured as an individual event giving rise to an individual line.

Regeneration

     After 12 weeks on selection, the tissue was removed
from callus maintenance edium with PPT and was placed on
regeneration medium. Regeneration medium is 0.25MS3S5BA (0.25
mg/1 2,4 D, 5 mg/1 BAP, MS salts, 3% sucrose) for 2 weeks
followed by subculture to MS3S medium for regeneration of
plants. After 4 to 10 weeks, plants were removed and put into
GA 7's. Our line KC65 0-6, which became the #176 BT event,
produced a total of 38 plants.

Assays

     All plants, as they became established in the GA7's,
were tested by the chlorophenol red (CR) test for resistance to
PPT as described in U.S. Patent Application 07/759,243, filed
September 13, 1991, the relevant portions of which are hereby
incorporated herein by reference. This assay utilizes a pH
sensitive indicator dye to show which cells are growing in the
presence of PPT. Cells which grow produce a pH change in the
media and turn the indicator yellow (from red). Plants
expressing the resistance gene to PPT are easily seen in this
test. (#176 = 8 positive/30 negative) Plants positive by the
CR test were assayed by PCR for the presence of the synthetic
BT gene. (#176 = 5 positive/2 negative/1 dead)

     Plants positive by PCR for the syn-BT gene were sent to
the phytotron. Once established in the phytotron, they were
characterized using insect bioassays and ELISA analysis.
Plants were insect bioassayed using a standard European Corn

147

Borer assay (described in Example 5A) in which small pieces of leaf of clipped from a plant and placed in a small petri dish with a number of ECB neonate larvae. Plants are typically assayed at a height of about 6 inches. Plants showing 100% mortality to ECB in this assay are characterized further. ELISA data are shown below. Positive plants are moved to the greenhouse.

Greenhouse/Fertility

　　　　Plant number #176-11 was pollinated with wild-type 6N615 pollen. One tassel ear and one ear shoot were produced. All of the embryos from the tassel ear (11) and 56 kernels from Ear 1 were rescued. 294 kernels remained on the ear and dried down naturally.

　　　　Pollen from #176-11 was outcrossed to various maize genotypes 5N984, 5NA89, and 3N961. Embryos have been rescued from all 3 outcrosses (5N984 = 45; 5NA89 = 30; 3N961 = 8). Most of the kernels remained on the ears on the plants in the greenhouse and were dried down naturally.

DNA was isolated from plant #176-11 using standard techniques and analysed by Southern blot analysis. It was found to contain sequences which hybridize with probes generated from the synthetic cryIA(b) gene and with a probe generated from the PAT gene. These results showed integration of these genes into the genome of maize.

Experiment KC-64

Production of transgenic maize plants expressing the synthetic cryIA(b) gene using a constitutive promoter.

Tissue

Immature maize embryos, approximately 1.5-2.5 mm in length, were excised from an ear of genotype 6N615 14-15 days after pollination. The mother plant was grown in the greenhouse. Before excision, the ear was surface sterilized with 20% Clorox for 20 minutes and rinse 3 times with sterile water. Individual embryos were plated scutellum side in a 2 cm square area, 36 embryos to a plate, on the callus initiation medium, 2DG4 + 5 chloramben medium (N6 major salts, B5 minor salts, MS iron, 2% sucrose, with 5 mg/l chloramben, 20 mg/l glucose, and 10 ml G4 additions Table 1) added after autoclaving.

TABLE 1 - G4 Additions

Ingredient	per liter medium	
Casein hydrolysate	0.5	gm
Proline	1.38	gm
Nicotinic acid	.2	mg
Pyridoxine-HCl	.2	mg
Thiamine-HCl	.5	mg
Choline-HCl	.1	mg
Riboflavin	.05	mg
Biotin	.1	mg
Folic acid	.05	mg
Ca pantothenate	.1	mg
p-aminobenzoic acid	.05	mg
B12	.136	µg

Bombardment

Tissue was bombarded using the PDS-1000He Biolistics device. The tissue was placed on the shelf 8 cm below the

149

stopping screen shelf. The tissue was shot one time with the
DNA/gold microcarrier solution, 10 µl dried onto the
macrocarrier. The stopping screen used was hand punched at
ABRU using 10x10 stainless steel mesh screen. Rupture discs of
1550 psi value were used. After bombardment, the embryos were
cultured in the dark at 25° C.

Preparation of DNA for delivery

The microcarrier was prepared essentially according to
the instructions supplied with the Biolistic device. While
vortexing 50 µl 1.0 µ gold microcarrier, added 3.2 µl pCIB4418
(0.85 µg/µl) [#905] + 2 µl pCIB3064   0.895 µg/µl) [#456] +
1.6 µl pCIB3007A (1.7 µg/µl) [#152] followed by 50 µl 2.5 M
$CaCl_2$, then 20 µl 0.1 M spermidine (free base, TC grade). The
resulting mixture was vortexed 3 minutes and microfuged for 10
sec. The supernatant was removed and the microcarriers washed
2 times with 250 µl of 100% EtOH (HPLC grade) by vortexing
briefly, centrifuging and removing the supernatant. The
microcarriers are resuspended in 65 µl 100% EtOH.

Callus formation

Embryos were transferred to callus initiation medium
with 3 mg/l PPT 1 day after bombardment. Embryos were scored
for callus initiation at 2 and 3 weeks after bombardment. Any
responses were transferred to callus maintenance medium, 2DG4 +
0.5 2,4-D medium with 3 mg/L PPT. Callus maintenance medium is
N6 major salts, B5 minor salts, MS iron, 2% sucrose, with 0.5

mg/l 2,4-D, 20 mg/l glucose, and 10 ml G4 additions added after

utoclaving. Embryogenic callus was subcultured every 2 weeks

to fresh maintenance medium containing 3 mg/L PPT. All callus

was incubated in the dark at 25°C.

The Type I callus formation response was 18%. Every

embryo which produced callus was cultured as an individual

event giving rise to an individual line.

Regeneration

After 12 weeks on selection, the tissue was removed

from callus maintenance medium with PPT and was placed on

regeneration medium and incubated at 25° C using a 16 hour

light (50 µE .m-2 . s-1) / 8 hour dark photoperiod.

Regeneration medium is 0.25MS3S5BA (0.25 mg/l 2,4 D, 5 mg/l

BAP, MS salts, 3% sucrose) for 2 weeks followed by subculture

to MS3S medium for regeneration of plants. After 4 to 10

weeks, plants were removed and put into GA 7's. Our line KC64

0-1, which became the #170 BT event, produced 55 plants. Our

line KC64 0-7, which became the #171 BT event, produced a total

of 33 plants.

Assays

Eleven plants, as they became established in the GA7's,

were tested by the chlorophenol red (CR) test for resistance to

PPT as per Shillito, et al, above. This assay utilizes a pH

sensitive indicator dye to show which cells are growing in the

presence of PPT. Cells which grow produce a pH change in the

media and turn the indicator yellow (from red). Plants expressing the resistance gene to PPT are easily seen in this test. Plants positive by the CR test were assayed by PCR for the presence of the synthetic BT gene. (Event 170 = 37 positive/18 negative; #171 = 25 positive/8 negative).

Plants positive by PCR for the syn-Bt gene were sent to the phytotron. Once established in the phytotron, they were characterized using insect bioassays and ELISA analysis. Plants were insect bioassayed using a standard European corn borer assay (see below) in which small pieces of leaf of clipped from a lant and placed in a small petri dish with a number of ECB neonate larvae. Plants are typically assayed at a height of about 6 inches. Plants showing 100% mortality to ECB in this assay are characterized further. ELISA data are shown below. Positive plants are moved to the greenhouse.

Basta screening

Eight of the mature plants from the #170 event were selected for evaluation of Basta [Hoechst] resistance. On one middle leaf per plant, an area approximately 10-14 cm long X the leaf width was painted with 0, 0.4, 1.0 or 2.0% (10 ml of 200 g/L diluted to 100 ml with deionized water) aqueous Basta containing 2 drops of Tween 20/100 ml. Two plants were tested per level. Eight wild-type 6N615 plants of the same approximate age were treated as controls. All plants were observed at 4 and 7 days. All of the control plants eventually died. Throughout the study, none of the #170 plants displayed

any damage due to the herbicide.

Pollination

All tassel ears, first ear and, if available, the second ear on the #170 and #171 plants were pollinated with wild-type 6N615 pollen. At least 90% of the plants were female fertile.

Pollen from #171 plants was outcrossed to genotypes 6N615, 5N984, 5NA89, 6F010, 5NA56, 2N217AF, 2ND01 and 3N961. At least 90% of the plants were shown to be male fertile.

Embryo Rescue

Embryos from the #171 event have been "rescued." Fourteen to 16 days after pollination, the ear tip with 25-50 kernels was cut from the ear with a coping saw. Prior to cutting, the husks were gently peeled away to expose the upper portion of the ear. The cut end of the ear on the plant was painted with Captan fungicide and the husks replaced. The seed remaining on the plant was allowed to dry naturally.

The excised ear piece was surface sterilized with 20% Clorox for 20 minutes and rinsed 3 times with sterile water. Individual embryos were excised and plated scutellum side up on B5 medium [Gamborg] containing 2% sucrose. B5 itamins are added to the medium after autoclaving. Four embryos were plated per GA7 container and the containers incubated in the dark. When germination occurred, the containers were moved to a light culture room and incubated at 25° C using a 16 hour

light (50 μE .m-2 . s-1) / 8 hour dark photoperiod. The
germination frequency is 94%.

Progeny from 15 plants of the #171 event and 2 of the
#176 event were rescued using standard embryo rescue techniques
and evaluated. All plants were evaluated by insect assay.
Plants from the #171 event were also tested in the
histochemical GUS assay. In both the insect assay and the GUS
assay, the ratio of segregation of the transgenes was 1:1, as
expected for a single locus insertion event.

EXAMPLE 48. Analysis of transgenic maize plants
             ELISA ASSAY

Detection of cryIA(b) gene expression in transgenic
maize is monitored using European corn borer(ECB) insect
bioassays and ELISA analysis for a quantitative determination
of the level of cryIA(b) protein obtained.
Quantitative determination of cryIA(b) IP in the leaves of
transgenic plants was performed using enzyme-linked
immunosorbant assays (ELISA) as disclosed in Clark M F, Lister
R M, Bar-Joseph M: ELISA Techniques. In: Weissbach A, Weissbach
H (eds) Methods in Enzymology 118:742-766, Academic Press,
Florida (1986). Immunoaffinity purified polyclonal rabbit and
goat antibodies specific for the B. thuringiensis subsp.
kurstaki IP were used to determine ng IP per mg soluble protein
from crude extracts of leaf samples. The sensitivity of the
double sandwich ELISA is 1-5 ng IP per mg soluble protein using
50 ug of total protein per ELISA microtiter dish well.

Corn extracts were made by grinding leaf tissue in gauze lined plastic bags using a hand held ball-bearing homogenizer (AGDIA, Elkart IN.) in the presence of extraction buffer (50 mM Na_2CO_3 pH 9.5, 100 mM NaCl, 0.05% Triton, 0.05% Tween, 1 mM PMSF and 1 μM leupeptin). Protein determination was performed using the Bio-Rad (Richmond, CA) protein assay.

Using the above procedure, the primary maize transformants described above were analyzed for the presence of cryIA(b) protein using ELISA. These plants varied in height from 6 inches to about three feet at the time of analysis.

Plant	Bt ng/mg soluble protein	5/27/91
176-8	0	0
176-10	700	1585
176-11	760	2195
171-4A	59	
171-6	50	
171-8	60	
171-9	280	
171-13	77	
171-14A	43	
171-14B	60	
171-15	55	
171-16A	13	
171-16B	19	
171-18	19	
176-30	1160	
171-32	980	
171-31	166	
171-30	370	
71-14		
#10 leaf	26	
1 leaf	17	
plant 171-16		
#9 leaf	40	
#1 leaf	120	

EUROPEAN CORN BORER ASSAY

1. One to four 4 cm sections are cut from an extended leaf of a corn plant.

2. Each leaf piece is placed on a moistened filter disc in a 50 X 9 mm petri dish.

3. Five neonate European corn borer larvae are placed on each leaf piece. (Making a total of 5-20 larvae per plant.)

4. The petri dishes are incubated at 29.5 °C.

5. Leaf feeding damage and mortality data are scored at 24, 48, and 72 hours.

EXAMPLE 49.   Expression of Bt endotoxin in progeny of transformed maize plants

The transformed maize plants were fully fertile and were crossed with several genotypes of maize. Progeny from these crosses were analyzed for their ability to kill European corn borer (ECB) in a standard ECB bioassay (described immediately above) as well as for the presence of the cryIA(b) protein using ELISA as described above. The ability to kill ECB and the production of cryIA(b) protein correlated. These traits segregated to the progeny with a 1:1 ratio, indicating a single site of insertion for the active copy of the synthetic gene. This 1:1 ratio was true for both the constitutive promoter/synthetic-cryIA(b) plants and the tissue specific promoter/synthetic-cryIA(b) plants (data not shown).

Fig. 23A is a table containing a small subset of the

total number of progeny analyzed. This table is representative
of a number of different crosses.

Insect assays were done with <u>Diatrea</u> <u>saccharalis</u> and
<u>Ostrinia</u> <u>nubilalis</u> using leaf material (as described above) of
transgenic progeny containing a maize optimized CryIA(b) gene.
The results of these assays are shown in Fig. 23B. They
demonstrate that the maize optimized CryIA(b) gene functions in
transformed maize to provide resistance to Sugarcane borer and
<u>Ostrinia</u> <u>nubilalis</u>.

<u>EXAMPLE 50.</u> EXPRESSION OF THE CRYIA(b) GENE IN MAIZE POLLEN

Progeny of the transformed maize plants containing the
chimeric pollen promoter/synthetic cryIA(b) gene derived from
pCIB4431 were grown in the field to maturity. Pollen was
collected and analyzed for the presence of the cryIA(b) protein
using standard ELISA techniquesd as described elsewhere. High
levels of cryIA(b) protein were detected in the pollen.
Progeny from the 35S promoter/synthetic cryIA(b) transformed
plant were grown in the greenhouse. Pollen from these plants
was analyzed using ELISA, and cryIA(b) protein was detected.

Results are shown below in Figure 23C.

It is recognized that factors including selection of
plant lines, plant genotypes, synthetic sequences and the like,
may also affect expression.

EXAMPLE 51. EXPRESSION OF THE CRYIA(b) GENE FUSED TO A
PITH-PREFERRED PROMOTER.

pCIB4433 (Fig. 36) is a plasmid containing the maize
optimized CryIA(b) gene fused with the pith-preferred promoter

isolated from maize. This plasmid was constructed using a three-way ligation consisting of:

1) pCIB4418, cut with BstEII and BamHI; 1.8 Kb fragment

2) pBtin1, cut with NsiI and BstEII; 5.9 Kb fragment; pBtin1 is described elsewhere in this application

3) PCR fragment VI-151 was generated in a PCR reaction using standard conditions as described elsewhere in this application.

PCR primers utilized were:

KE150A28: 5' -ATT CGC ATG CAT GTT TCA TTA TC-3'

KE151A28: 5' - GCT GGT ACC ACG GAT CCG TCG CTT CTG TGC AAC AAC C-3'

After the PCR reaction, the DNA was checked on an agarose gel to make sure the reaction had proceeded properly. DNA was recovered from the PCR reaction using standard conditions described elsewhere and subsequently cut with the restriction enzymes NsiI and BamHI using standard condition. After cutting, the fragment was run on a 2% NuSieve gel and the desired band recovered as described elsewhere. The DNA was used in the ligation described above.

After ligation (under standard condition), the DNA was transformed into competent E. coli cell.

Transformation was carried out using microprojectile bombardment essentially as described elsewhere in this application. Embryos were transferred to medium containing 10∂µg/ml PPT 24 hours after microprojectile bombardment. Resulting callus was transferred to medium containing 40 µg/ml

PPT after four weeks. Plants were regenerated without

selection.

A small sample of plants (3-5) was assayed by PCR for

each event. Further codes were added to indicate different

positions and distances of embryos with respect to the

microprojectile bombardment device. Plants were sent to the

greenhouse having the following codes:

<table>
<tr><td>JS21A TOP</td><td>Plants <u>B.t.</u> PCR Positive</td></tr>
<tr><td>JS21A MID</td><td>Plants <u>B.t.</u> PCR Positive</td></tr>
<tr><td>JS21C BOT</td><td>Plants <u>B.t.</u> PCR Positive</td></tr>
<tr><td>JS22D MID</td><td>Plants <u>B.t.</u> PCR Positive</td></tr>
<tr><td>JS23B MID</td><td>Plants <u>B.t.</u> PCR Negative (for control)</td></tr>
</table>

Leaf samples from the regenerated plants were

bioassayed for insecticidal activity against European corn

borer as described in Example 48 with the results shown in Fig.

23D.

ELISA analysis of leaf samples to quantify the level of

CryIA(b) protein expressed in the leaves was carried out as

described in Example 48 with the results shown in Fig. 24E.

<u>Deposits</u>

The following plasmids have been deposited with the

Agricultural Research Culture Collection (NRRL) (1818 N.

University St., Peoria, IL 61604) under the provisions of the

Budapest Treaty: pCIB4418, pCIB4420, pCIB4429, pCIB4431,

pCIB4433, pCIB5601, pCIB3166 and pCIB3171.

* * * * * * * * * *

159

The present invention has been described with reference to specific embodiments thereof; however it will be appreciated that numerous variations, modifications, and embodiments are possible. Accordingly, all such variations, modifications and embodiments are to be regarded as being within the spirit and scope of the present invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT: Evola, Stephen V.
                Crossland, Lyle D.
                Wright, Martha S.
                Merlin, Ellis J.
                Launis, Karen L.
                Rothstein, Steven J.

   (ii) TITLE OF INVENTION: SYNTHETIC DNA SEQUENCE HAVING ENHANCED
         INSECTICIDAL ACTIVITY IN MAIZE

  (iii) NUMBER OF SEQUENCES: 5

   (iv) CORRESPONDENCE ADDRESS:
        (A) ADDRESSEE: CIBA-GEIGY Corporation
        (B) STREET: 7 Skyline Drive
        (C) CITY: Hawthorne
        (D) STATE: New York
        (E) COUNTRY: USA
        (F) ZIP: 10532

    (v) COMPUTER READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version #1.25

   (vi) CURRENT APPLICATION DATA:
        (A) APPLICATION NUMBER: US
        (B) FILING DATE:
        (C) CLASSIFICATION:

  (viii) ATTORNEY/AGENT INFORMATION:
        (A) NAME:
        (B) REGISTRATION NUMBER:
        (C) REFERENCE/DOCKET NUMBER: CGC 1577/CIP

   (ix) TELECOMMUNICATION INFORMATION:
        (A) TELEPHONE: (919)541-8615
        (B) TELEFAX: (919)541-8689

(2) INFORMATION FOR SEQ ID NO:1:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4360 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
    (A) ORGANISM: Bacillus thuringiensis
    (B) STRAIN: Kurstaki HD-1
    (C) INDIVIDUAL ISOLATE: CryIA(b) insecticidal crystal
        protein gene

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GTTAACACCC TGGGTCAAAA ATTGATATTT AGTAAAATTA GTTGCACTTT GTGCATTTTT      60

TCATAAGATG AGTCATATGT TTTAAATTGT AGTAATGAAA AACAGTATTA TATCATAATG     120

AATTGGTATC TTAATAAAAG AGATGGAGGT AACTTATGGA TAACAATCCG AACATCAATG     180

AATGCATTCC TTATAATTGT TTAAGTAACC CTGAAGTAGA AGTATTAGGT GGAGAAAGAA     240

TAGAAACTGG TTACACCCCA ATCGATATTT CCTTGTCGCT AACGCAATTT CTTTTGAGTG     300

AATTTGTTCC CGGTGCTGGA TTTGTGTTAG GACTAGTTGA TATAATATGG GGAATTTTTG     360

GTCCCTCTCA ATGGGACGCA TTTCTTGTAC AAATTGAACA GTTAATTAAC CAAAGAATAG     420

AAGAATTCGC TAGGAACCAA GCCATTTCTA GATTAGAAGG ACTAAGCAAT CTTTATCAAA     480

TTTACGCAGA ATCTTTTAGA GAGTGGGAAG CAGATCCTAC TAATCCAGCA TTAAGAGAAG     540

AGATGCGTAT TCAATTCAAT GACATGAACA GTGCCCTTAC AACCGCTATT CCTCTTTTTG     600

CAGTTCAAAA TTATCAAGTT CCTCTTTTAT CAGTATATGT TCAAGCTGCA AATTTACATT     660

TATCAGTTTT GAGAGATGTT TCAGTGTTTG GACAAAGGTG GGGATTTGAT GCCGCGACTA     720

TCAATAGTCG TTATAATGAT TTAACTAGGC TTATTGGCAA CTATACAGAT CATGCTGTAC     780

GCTGGTACAA TACGGGATTA GAGCGTGTAT GGGGACCGGA TTCTAGAGAT TGGATAAGAT     840

ATAATCAATT TAGAAGAGAA TTAACACTAA CTGTATTAGA TATCGTTTCT CTATTTCCGA     900

ACTATGATAG TAGAACGTAT CCAATTCGAA CAGTTTCCCA ATTAACAAGA GAAATTTATA     960

CAAACCCAGT ATTAGAAAAT TTTGATGGTA GTTTTCGAGG CTCGGCTCAG GGCATAGAAG    1020

GAAGTATTAG GAGTCCACAT TTGATGGATA TACTTAACAG TATAACCATC TATACGGATG    1080

CTCATAGAGG AGAATATTAT TGGTCAGGGC ATCAAATAAT GGCTTCTCCT GTAGGGTTTT    1140
```

162

```
CGGGGCCAGA ATTCACTTTT CCGCTATATG GAACTATGGG AAATGCAGCT CCACAACAAC    1200

GTATTGTTGC TCAACTAGGT CAGGGCGTGT ATAGAACATT ATCGTCCACT TTATATAGAA    1260

GACCTTTTAA TATAGGGATA AATAATCAAC AACTATCTGT TCTTGACGGG ACAGAATTTG    1320

CTTATGGAAC CTCCTCAAAT TTGCCATCCG CTGTATACAG AAAAAGCGGA ACGGTAGATT    1380

CGCTGGATGA AATACCGCCA CAGAATAACA ACGTGCCACC TAGGCAAGGA TTTAGTCATC    1440

GATTAAGCCA TGTTTCAATG TTTCGTTCAG GCTTTAGTAA TAGTAGTGTA AGTATAATAA    1500

GAGCTCCTAT GTTCTCTTGG ATACATCGTA GTGCTGAATT TAATAATATA ATTCCTTCAT    1560

CACAAATTAC ACAAATACCT TTAACAAAAT CTACTAATCT TGGCTCTGGA ACTTCTGTCG    1620

TTAAAGGACC AGGATTTACA GGAGGAGATA TTCTTCGAAG AACTTCACCT GGCCAGATTT    1680

CAACCTTAAG AGTAAATATT ACTGCACCAT TATCACAAAG ATATCGGGTA AGAATTCGCT    1740

ACGCTTCTAC CACAAATTTA CAATTCCATA CATCAATTGA CGGAAGACCT ATTAATCAGG    1800

GGAATTTTTC AGCAACTATG AGTAGTGGGA GTAATTTACA GTCCGGAAGC TTTAGGACTG    1860

TAGGTTTTAC TACTCCGTTT AACTTTTCAA ATGGATCAAG TGTATTTACG TTAAGTGCTC    1920

ATGTCTTCAA TTCAGGCAAT GAAGTTTATA TAGATCGAAT TGAATTTGTT CCGGCAGAAG    1980

TAACCTTTGA GGCAGAATAT GATTTAGAAA GAGCACAAAA GGCGGTGAAT GAGCTGTTTA    2040

CTTCTTCCAA TCAAATCGGG TTAAAAACAG ATGTGACGGA TTATCATATT GATCAAGTAT    2100

CCAATTTAGT TGAGTGTTTA TCTGATGAAT TTTGTCTGGA TGAAAAAAAA GAATTGTCCG    2160

AGAAAGTCAA ACATGCGAAG CGACTTAGTG ATGAGCGGAA TTTACTTCAA GATCCAAACT    2220

TTAGAGGGAT CAATAGACAA CTAGACCGTG GCTGGAGAGG AAGTACGGAT ATTACCATCC    2280

AAGGAGGCGA TGACGTATTC AAAGAGAATT ACGTTACGCT ATTGGGTACC TTTGATGAGT    2340

GCTATCCAAC GTATTTATAT CAAAAAATAG ATGAGTCGAA ATTAAAAGCC TATACCCGTT    2400

ACCAATTAAG AGGGTATATC GAAGATAGTC AAGACTTAGA AATCTATTTA ATTCGCTACA    2460

ATGCCAAACA CGAAACAGTA AATGTGCCAG GTACGGGTTC CTTATGGCCG CTTTCAGCCC    2520

CAAGTCCAAT CGGAAAATGT GCCCATCATT CCCATCATTT CTCCTTGGAC ATTGATGTTG    2580

GATGTACAGA CTTAAATGAG GACTTAGGTG TATGGGTGAT ATTCAAGATT AAGACGCAAG    2640

ATGGCCATGC AAGACTAGGA AATCTAGAAT TTCTCGAAGA GAAACCATTA GTAGGAGAAG    2700

CACTAGCTCG TGTGAAAAGA GCGGAGAAAA AATGGAGAGA CAAACGTGAA AAATTGGAAT    2760
```

163

```
GGGAAACAAA TATTGTTTAT AAAGAGGCAA AAGAATCTGT AGATGCTTTA TTTGTAAACT    2820

CTCAATATGA TAGATTACAA GCGGATACCA ACATCGCGAT GATTCATGCG GCAGATAAAC    2880

GCGTTCATAG CATTCGAGAA GCTTATCTGC CTGAGCTGTC TGTGATTCCG GGTGTCAATG    2940

CGGCTATTTT TGAAGAATTA GAAGGGCGTA TTTTCACTGC ATTCTCCCTA TATGATGCGA    3000

GAAATGTCAT TAAAAATGGT GATTTTAATA ATGGCTTATC CTGCTGGAAC GTGAAAGGGC    3060

ATGTAGATGT AGAAGAACAA AACAACCACC GTTCGGTCCT TGTTGTTCCG GAATGGGAAG    3120

CAGAAGTGTC ACAAGAAGTT CGTGTCTGTC CGGGTCGTGG CTATATCCTT CGTGTCACAG    3180

CGTACAAGGA GGGATATGGA GAAGGTTGCG TAACCATTCA TGAGATCGAG AACAATACAG    3240

ACGAACTGAA GTTTAGCAAC TGTGTAGAAG AGGAAGTATA TCCAAACAAC ACGGTAACGT    3300

GTAATGATTA TACTGCGACT CAAGAAGAAT ATGAGGGTAC GTACACTTCT CGTAATCGAG    3360

GATATGACGG AGCCTATGAA AGCAATTCTT CTGTACCAGC TGATTATGCA TCAGCCTATG    3420

AAGAAAAAGC ATATACAGAT GGACGAAGAG ACAATCCTTG TGAATCTAAC AGAGGATATG    3480

GGGATTACAC ACCACTACCA GCTGGCTATG TGACAAAAGA ATTAGAGTAC TTCCCAGAAA    3540

CCGATAAGGT ATGGATTGAG ATCGGAGAAA CGGAAGGAAC ATTCATCGTG GACAGCGTGG    3600

AATTACTTCT TATGGAGGAA TAATATATGC TTTATAATGT AAGGTGTGCA AATAAAGAAT    3660

GATTACTGAC TTGTATTGAC AGATAAATAA GGAAATTTTT ATATGAATAA AAAACGGGCA    3720

TCACTCTTAA AAGAATGATG TCCGTTTTTT GTATGATTTA ACGAGTGATA TTTAAATGTT    3780

TTTTTTGCGA AGGCTTTACT TAACGGGGTA CCGCCACATG CCCATCAACT TAAGAATTTG    3840

CACTACCCCC AAGTGTCAAA AAACGTTATT CTTTCTAAAA AGCTAGCTAG AAAGGATGAC    3900

ATTTTTTATG AATCTTTCAA TTCAAGATGA ATTACAACTA TTTTCTGAAG AGCTGTATCG    3960

TCATTTAACC CCTTCTCTTT TGGAAGAACT CGCTAAAGAA TTAGGTTTTG TAAAAGAAA    4020

ACGAAAGTTT TCAGGAAATG AATTAGCTAC CATATGTATC TGGGGCAGTC AACGTACAGC    4080

GAGTGATTCT CTCGTTCGAC TATGCAGTCA ATTACACGCC GCCACAGCAC TCTTATGAGT    4140

CCAGAAGGAC TCAATAAACG CTTTGATAAA AAAGCGGTTG AATTTTTGAA ATATATTTTT    4200

TCTGCATTAT GGAAAAGTAA ACTTTGTAAA ACATCAGCCA TTTCAAGTGC AGCACTCACG    4260

TATTTTCAAC GAATCCGTAT TTTAGATGCG ACGATTTTCC AAGTACCGAA ACATTTAGCA    4320

CATGTATATC CTGGGTCAGG TGGTTGTGCA CAAACTGCAG                          4360
```

(2) INFORMATION FOR SEQ ID NO:2:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3474 base pairs
        (B) TYPE: nucleic acid
        (C) STRANDEDNESS: single
        (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA

    (iii) HYPOTHETICAL: YES

    (iv) ANTI-SENSE: NO

    (vi) ORIGINAL SOURCE:
        (A) ORGANISM: Pure maize optimized synthetic BT CryIA(b)
           gene

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGAG    60

GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCCTG   120

AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCCTG   180

GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGATC   240

GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCCTG   300

GAGGGCCTGA GCAACCTGTA CCAGATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGAC   360

CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGCC   420

CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGTG   480

TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTGAGCGT GTTCGGCCAG   540

CGCTGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGATC   600

GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG GCCTGGAGCG CGTGTGGGGC   660

CCCGACAGCC GCGACTGGAT CCGCTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGTG   720

CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGTG   780

AGCCAGCTGA CCCGCGAGAT CTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC   840

CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG   900

AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG   960
```

```
ATCATGGCCA GCCCCGTGGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC    1020
ATGGGCAACG CCGCCCCCCA GCAGCGCATC GTGGCCCAGC TGGGCCAGGG CGTGTACCGC    1080
ACCCTGAGCA GCACCCTGTA CCGCCGCCCC TTCAACATCG GCATCAACAA CCAGCAGCTG    1140
AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG    1200
TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCCCAGAA CAACAACGTG    1260
CCCCCCCGCC AGGGCTTCAG CCACCGCCTG AGCCACGTGA GCATGTTCCG CAGCGGCTTC    1320
AGCAACAGCA GCGTGAGCAT CATCCGCGCC CCCATGTTCA GCTGGATCCA CCGCAGCGCC    1380
GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC    1440
AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG    1500
CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC    1560
CAGCGCTACC GCGTGCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC    1620
ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC    1680
CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC    1740
AGCAGCGTGT TCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC    1800
CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCG AGTACGACCT GGAGCGCGCC    1860
CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG    1920
ACCGACTACC ACATCGACCA GGTGAGCAAC CTGGTGGAGT GCCTGAGCGA CGAGTTCTGC    1980
CTGGACGAGA GAAGGAGCT GAGCGAGAAG GTGAAGCACG CCAAGCGCCT GAGCGACGAG    2040
CGCAACCTGC TGCAGGACCC CAACTTCCGC GGCATCAACC GCCAGCTGGA CCGCGGCTGG    2100
CGCGGCAGCA CCGACATCAC CATCCAGGGC GGCGACGACG TGTTCAAGGA GAACTACGTG    2160
ACCCTGCTGG CACCTTCGA CGAGTGCTAC CCCACCTACC TGTACCAGAA GATCGACGAG    2220
AGCAAGCTGA AGGCCTACAC CCGCTACCAG CTGCGCGGCT ACATCGAGGA CAGCCAGGAC    2280
CTGGAGATCT ACCTGATCCG CTACAACGCC AAGCACGAGA CCGTGAACGT.GCCCGGCACC    2340
GGCAGCCTGT GGCCCCTGAG CGCCCCCAGC CCCATCGGCA AGTGCGCCCA CCACAGCCAC    2400
CACTTCAGCC TGGACATCGA CGTGGGCTGC ACCGACCTGA ACGAGGACCT GGGCGTGTGG    2460
GTGATCTTCA AGATCAAGAC CCAGGACGGC CACGCCCGCC TGGGCAACCT GGAGTTCCTG    2520
GAGGAGAAGC CCCTGGTGGG CGAGGCCCTG GCCCGCGTGA AGCGCGCCGA GAAGAAGTGG    2580
```

166

```
CGCGACAAGC GCGAGAAGCT GGAGTGGGAG ACCAACATCG TGTACAAGGA GGCCAAGGAG        2640

AGCGTGGACG CCCTGTTCGT GAACAGCCAG TACGACCGCC TGCAGGCCGA CACCAACATC        2700

GCCATGATCC ACGCCGCCGA CAAGCGCGTG CACAGCATCC GCGAGGCCTA CCTGCCCGAG        2760

CTGAGCGTGA TCCCCGGCGT GAACGCCGCC ATCTTCGAGG AGCTGGAGGG CCGCATCTTC        2820

ACCGCCTTCA GCCTGTACGA CGCCCGCAAC GTGATCAAGA ACGGCGACTT CAACAACGGC        2880

CTGAGCTGCT GGAACGTGAA GGGCCACGTG GACGTGGAGG AGCAGAACAA CCACCGCAGC        2940

GTGCTGGTGG TGCCCGAGTG GGAGGCCGAG GTGAGCCAGG AGGTGCGCGT GTGCCCCGGC        3000

CGCGGCTACA TCCTGCGCGT GACCGCCTAC AAGGAGGGCT ACGGCGAGGG CTGCGTGACC        3060

ATCCACGAGA TCGAGAACAA CACCGACGAG CTGAAGTTCA GCAACTGCGT GGAGGAGGAG        3120

GTGTACCCCA ACAACACCGT GACCTGCAAC GACTACACCG CCACCCAGGA GGAGTACGAG        3180

GGCACCTACA CCAGCCGCAA CCGCGGCTAC GACGGCGCCT ACGAGAGCAA CAGCAGCGTG        3240

CCCGCCGACT ACGCCAGCGC CTACGAGGAG AAGGCCTACA CCGACGGCCG CCGCGACAAC        3300

CCCTGCGAGA GCAACCGCGG CTACGGCGAC TACACCCCCC TGCCCGCCGG CTACGTGACC        3360

AAGGAGCTGG AGTACTTCCC CGAGACCGAC AAGGTGTGGA TCGAGATCGG CGAGACCGAG        3420

GGCACCTTCA TCGTGGACAG CGTGGAGCTG CTGCTGATGG AGGAGTAGTA CATG             3474
```

(2) INFORMATION FOR SEQ ID NO:3:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 1961 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA

    (iii) HYPOTHETICAL: NO

     (iv) ANTI-SENSE: NO

     (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Truncated synthetic maize optimized BT
             CryIA(b) gene

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
GATCCAACAA TGGACAACAA CCCCAACATC AACGAGTGCA TCCCCTACAA CTGCCTGAGC        60
```

```
AACCCCGAGG TGGAGGTGCT GGGCGGCGAG CGCATCGAGA CCGGCTACAC CCCCATCGAC      120

ATCAGCCTGA GCCTGACCCA GTTCCTGCTG AGCGAGTTCG TGCCCGGCGC CGGCTTCGTG      180

CTGGGCCTGG TGGACATCAT CTGGGGCATC TTCGGCCCCA GCCAGTGGGA CGCCTTCCTG      240

GTGCAGATCG AGCAGCTGAT CAACCAGCGC ATCGAGGAGT TCGCCCGCAA CCAGGCCATC      300

AGCCGCCTGG AGGGCCTGAG CAACCTGTAC CAAATCTACG CCGAGAGCTT CCGCGAGTGG      360

GAGGCCGACC CCACCAACCC CGCCCTGCGC GAGGAGATGC GCATCCAGTT CAACGACATG      420

AACAGCGCCC TGACCACCGC CATCCCCCTG TTCGCCGTGC AGAACTACCA GGTGCCCCTG      480

CTGAGCGTGT ACGTGCAGGC CGCCAACCTG CACCTGAGCG TGCTGCGCGA CGTCAGCGTG      540

TTCGGCCAGC GCTGGGGCTT CGACGCCGCC ACCATCAACA GCCGCTACAA CGACCTGACC      600

CGCCTGATCG GCAACTACAC CGACCACGCC GTGCGCTGGT ACAACACCGG CCTGGAGCGC      660

GTGTGGGGTC CCGACAGCCG CGACTGGATC AGGTACAACC AGTTCCGCCG CGAGCTGACC      720

CTGACCGTGC TGGACATCGT GAGCCTGTTC CCCAACTACG ACAGCCGCAC CTACCCCATC      780

CGCACCGTGA GCCAGCTGAC CCGCGAGATT TACACCAACC CCGTGCTGGA GAACTTCGAC      840

GGCAGCTTCC GCGGCAGCGC CCAGGGCATC GAGGGCAGCA TCCGCAGCCC CCACCTGATG      900

GACATCCTGA ACAGCATCAC CATCTACACC GACGCCCACC GCGGCGAGTA CTACTGGAGC      960

GGCCACCAGA TCATGGCCAG CCCCGTCGGC TTCAGCGGCC CCGAGTTCAC CTTCCCCCTG     1020

TACGGCACCA TGGGCAACGC TGCACCTCAG CAGCGCATCG TGGCACAGCT GGGCCAGGGA     1080

GTGTACCGCA CCCTGAGCAG CACCCTGTAC CGTCGACCTT CAACATCGG CATCAACAAC     1140

CAGCAGCTGA GCGTGCTGGA CGGCACCGAG TTCGCCTACG GCACCAGCAG CAACCTGCCC     1200

AGCGCCGTGT ACCGCAAGAG CGGCACCGTG GACAGCCTGG ACGAGATCCC CCCTCAGAAC     1260

AACAACGTGC CACCTCGACA GGGCTTCAGC CACCGTCTGA GCCACGTGAG CATGTTCCGC     1320

AGTGGCTTCA GCAACAGCAG CGTGAGCATC ATCCGTGCAC CTATGTTCAG CTGGATTCAC     1380

CGCAGTGCCG AGTTCAACAA CATCATCCCC AGCAGCCAGA TCACCCAGAT CCCCCTGACC     1440

AAGAGCACCA ACCTGGGCAG CGGCACCAGC GTGGTGAAGG CCCCGGCTT CACCGGCGGC     1500

GACATCCTGC GCCGCACCAG CCCCGGCCAG ATCAGCACCC TGCGCGTGAA CATCACCGCC     1560

CCCCTGAGCC AGCGCTACCG CGTCCGCATC CGCTACGCCA GCACCACCAA CCTGCAGTTC     1620

CACACCAGCA TCGACGGCCG CCCCATCAAC CAGGGCAACT TCAGCGCCAC CATGAGCAGC     1680
```

168

```
GGCAGCAACC TGCAGAGCGG CAGCTTCCGC ACCGTGGGCT TCACCACCCC CTTCAACTTC    1740

AGCAACGGCA GCAGCGTGTT CACCCTGAGC GCCCACGTGT TCAACAGCGG CAACGAGGTG    1800

TACATCGACC GCATCGAGTT CGTGCCCGCC GAGGTGACCT TCGAGGCCGA GTACGACCTG    1860

GAGAGGGCTC AGAAGGCCGT GAACGAGCTG TTCACCAGCA GCAACCAGAT CGGCCTGAAG    1920

ACCGACGTGA CCGACTACCA CATCGATCAG GTGTAGGAGC T                        1961
```

(2) INFORMATION FOR SEQ ID NO:4:

      (i) SEQUENCE CHARACTERISTICS:
          (A) LENGTH: 3508 base pairs
          (B) TYPE: nucleic acid
          (C) STRANDEDNESS: single
          (D) TOPOLOGY: linear

     (ii) MOLECULE TYPE: DNA

     (iv) ANTI-SENSE: NO

     (vi) ORIGINAL SOURCE:
          (A) ORGANISM: Full length synthetic maize optimized BT
              CryIA(b) gene

     (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
GATCCAACAA TGGACAACAA CCCCAACATC AACGAGTGCA TCCCCTACAA CTGCCTGAGC     60

AACCCCGAGG TGGAGGTGCT GGGCGGCGAG CGCATCGAGA CCGGCTACAC CCCCATCGAC    120

ATCAGCCTGA GCCTGACCCA GTTCCTGCTG AGCGAGTTCG TGCCCGGCGC CGGCTTCGTG    180

CTGGGCCTGG TGGACATCAT CTGGGGCATC TTCGGCCCCA GCCAGTGGGA CGCCTTCCTG    240

GTGCAGATCG AGCAGCTGAT CAACCAGCGC ATCGAGGAGT TCGCCCGCAA CCAGGCCATC    300

AGCCGCCTGG AGGGCCTGAG CAACCTGTAC CAAATCTACG CCGAGAGCTT CCGCGAGTGG    360

GAGGCCGACC CCACCAACCC CGCCCTGCGC GAGGAGATGC GCATCCAGTT CAACGACATG    420

AACAGCGCCC TGACCACCGC CATCCCCCTG TTCGCCGTGC AGAACTACCA GGTGCCCCTG    480

CTGAGCGTGT ACGTGCAGGC CGCCAACCTG CACCTGAGCG TGCTGCGCGA CGTCAGCGTG    540

TTCGGCCAGC GCTGGGGCTT CGACGCCGCC ACCATCAACA GCCGCTACAA CGACCTGACC    600

CGCCTGATCG GCAACTACAC CGACCACGCC GTGCGCTGGT ACAACACCGG CCTGGAGCGC    660

GTGTGGGGTC CCGACAGCCG CGACTGGATC AGGTACAACC AGTTCCGCCG CGAGCTGACC    720
```

```
CTGACCGTGC TGGACATCGT GAGCCTGTTC CCCAACTACG ACAGCCGCAC CTACCCCATC    780
CGCACCGTGA GCCAGCTGAC CCGCGAGATT TACACCAACC CCGTGCTGGA GAACTTCGAC    840
GGCAGCTTCC GCGGCAGCGC CCAGGGCATC GAGGGCAGCA TCCGCAGCCC CCACCTGATG    900
GACATCCTGA ACAGCATCAC CATCTACACC GACGCCCACC GCGGCGAGTA CTACTGGAGC    960
GGCCACCAGA TCATGGCCAG CCCCGTCGGC TTCAGCGGCC CCGAGTTCAC CTTCCCCCTG   1020
TACGGCACCA TGGGCAACGC TGCACCTCAG CAGCGCATCG TGGCACAGCT GGGCCAGGGA   1080
GTGTACCGCA CCCTGAGCAG CACCCTGTAC CGTCGACCTT TCAACATCGG CATCAACAAC   1140
CAGCAGCTGA GCGTGCTGGA CGGCACCGAG TTCGCCTACG GCACCAGCAG CAACCTGCCC   1200
AGCGCCGTGT ACCGCAAGAG CGGCACCGTG GACAGCCTGG ACGAGATCCC CCCTCAGAAC   1260
AACAACGTGC CACCTCGACA GGGCTTCAGC CACCGTCTGA GCCACGTGAG CATGTTCCGC   1320
AGTGGCTTCA GCAACAGCAG CGTGAGCATC ATCCGTGCAC CTATGTTCAG CTGGATTCAC   1380
CGCAGTGCCG AGTTCAACAA CATCATCCCC AGCAGCCAGA TCACCCAGAT CCCCCTGACC   1440
AAGAGCACCA ACCTGGGCAG CGGCACCAGC GTGGTGAAGG CCCCGGCTT CACCGGCGGC   1500
GACATCCTGC GCCGCACCAG CCCCGGCCAG ATCAGCACCC TGCGCGTGAA CATCACCGCC   1560
CCCCTGAGCC AGCGCTACCG CGTCCGCATC CGCTACGCCA GCACCACCAA CCTGCAGTTC   1620
CACACCAGCA TCGACGGCCG CCCCATCAAC CAGGGCAACT TCAGCGCCAC CATGAGCAGC   1680
GGCAGCAACC TGCAGAGCGG CAGCTTCCGC ACCGTGGGCT TCACCACCCC CTTCAACTTC   1740
AGCAACGGCA GCAGCGTGTT CACCCTGAGC GCCCACGTGT CAACAGCGG CAACGAGGTG   1800
TACATCGACC GCATCGAGTT CGTGCCCGCC GAGGTGACCT TCGAGGCCGA GTACGACCTG   1860
GAGAGGGCTC AGAAGGCCGT GAACGAGCTG TTCACCAGCA GCAACCAGAT CGGCCTGAAG   1920
ACCGACGTGA CCGACTACCA CATCGATCAG GTGAGCAACC TGGTGGAGTG CCTGAGCGAC   1980
GAGTTCTGCC TGGACGAGAA GAAGGAGCTG AGCGAGAAGG TGAAGCACGC CAAGCGCCTG   2040
AGCGACGAGC GCAACCTGCT GCAGGACCCC AACTTCCGCG GCATCAACCG CCAGCTGGAC   2100
CGCGGCTGGC GCGGCAGCAC CGACATCACC ATCCAGGGCG GCGACGACGT GTTCAAGGAG   2160
AACTACGTGA CCCTGCTGGG CACCTTCGAC GAGTGCTACC CCACCTACCT GTACCAGAAG   2220
ATCGACGAGA GCAAGCTGAA GGCCTACACC CGCTACCAGC TGCGCGGCTA CATCGAGGAC   2280
AGCCAGGACC TGGAGATCTA CCTGATCCGC TACAACGCCA AGCACGAGAC CGTGAACGTG   2340
```

```
CCCGGCACCG GCAGCCTGTG GCCCCTGAGC GCCCCCAGCC CCATCGGCAA GTGCGCCCAC     2400

CACAGCCACC ACTTCAGCCT GGACATCGAC GTGGGCTGCA CCGACCTGAA CGAGGACCTG     2460

GGCGTGTGGG TGATCTTCAA GATCAAGACC CAGGACGGCC ACGCCCGCCT GGGCAACCTG     2520

GAGTTCCTGG AGGAGAAGCC CCTGGTGGGC GAGGCCCTGG CCCGCGTGAA GCGCGCCGAG     2580

AAGAAGTGGC GCGACAAGCG CGAGAAGCTG GAGTGGGAGA CCAACATCGT GTACAAGGAG     2640

GCCAAGGAGA GCGTGGACGC CCTGTTCGTG AACAGCCAGT ACGACCGCCT GCAGGCCGAC     2700

ACCAACATCG CCATGATCCA CGCCGCCGAC AAGCGCGTGC ACAGCATTCG CGAGGCCTAC     2760

CTGCCCGAGC TGAGCGTGAT CCCCGGCGTG AACGCCGCCA TCTTCGAGGA GCTGGAGGGC     2820

CGCATCTTCA CCGCCTTCAG CCTGTACGAC GCCCGCAACG TGATCAAGAA CGGCGACTTC     2880

AACAACGGCC TGAGCTGCTG GAACGTGAAG GGCCACGTGG ACGTGGAGGA GCAGAACAAC     2940

CACCGCAGCG TGCTGGTGGT GCCCGAGTGG GAGGCCGAGG TGAGCCAGGA GGTGCGCGTG     3000

TGCCCCGGCC GCGGCTACAT CCTGCGCGTG ACCGCCTACA AGGAGGGCTA CGGCGAGGGC     3060

TGCGTGACCA TCCACGAGAT CGAGAACAAC ACCGACGAGC TCAAGTTCAG CAACTGCGTG     3120

GAGGAGGAGG TGTACCCCAA CAACACCGTG ACCTGCAACG ACTACACCGC CACCCAGGAG     3180

GAGTACGAGG GCACCTACAC CAGCCGCAAC CGCGGCTACG ACGGCGCCTA CGAGAGCAAC     3240

AGCAGCGTGC CCGCCGACTA CGCCAGCGCC TACGAGGAGA AGGCCTACAC CGACGGCCGC     3300

CGCGACAACC CCTGCGAGAG CAACCGCGGC TACGGCGACT ACACCCCCCT GCCCGCCGGC     3360

TACGTGACCA AGGAGCTGGA GTACTTCCCC GAGACCGACA AGGTGTGGAT CGAGATCGGC     3420

GAGACCGAGG GCACCTTCAT CGTGGACAGC GTGGAGCTGC TGCTGATGGA GGAGTAGTAC     3480

ATGTGATAGT ACGTAAGCTC GAGGATCT                                       3508
```

(2) INFORMATION FOR SEQ ID NO:5:

    (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 1845 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear

    (ii) MOLECULE TYPE: DNA

    (iv) ANTI-SENSE: NO

    (vi) ORIGINAL SOURCE:
      (A) ORGANISM: Truncated synthetic BT gene (Perlak et al.)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATGGACAACA ACCCAAACAT CAACGAATGC ATTCCATACA ACTGCTTGAG TAACCCAGAA    60
GTTGAAGTAC TTGGTGGAGA ACGCATTGAA ACCGGTTACA CTCCCATCGA CATCTCCTTG   120
TCCTTGACAC AGTTTCTGCT CAGCGAGTTC GTGCCAGGTG CTGGGTTCGT TCTCGGACTA   180
GTTGACATCA TCTGGGGTAT CTTTGGTCCA TCTCAATGGG ATGCATTCCT GGTGCAAATT   240
GAGCAGTTGA TCAACCAGAG GATCGAAGAG TTCGCCAGGA ACCAGGCCAT CTCTAGGTTG   300
GAAGGATTGA GCAATCTCTA CCAAATCTAT GCAGAGAGCT TCAGAGAGTG GGAAGCCGAT   360
CCTACTAACC CAGCTCTCCG CGAGGAAATG CGTATTCAAT TCAACGACAT GAACAGCGCC   420
TTGACCACAG CTATCCCATT GTTCGCAGTC CAGAACTACC AAGTTCCTCT CTTGTCCGTG   480
TACGTTCAAG CAGCTAATCT TCACCTCAGC GTGCTTCGAG ACGTTAGCGT GTTTGGGCAA   540
AGGTGGGGAT TCGATGCTGC AACCATCAAT AGCCGTTACA ACGACCTTAC TAGGCTGATT   600
GGAAACTACA CCGACCACGC TGTTCGTTGG TACAACACTG GCTTGGAGCG TGTCTGGGGT   660
CCTGATTCTA GAGATTGGAT TAGATACAAC CAGTTCAGGA GAGAATTGAC CCTCACAGTT   720
TTGGACATTG TGTCTCTCTT CCCGAACTAT GACTCCAGAA CCTACCCTAT CCGTACAGTG   780
TCCCAACTTA CCAGAGAAAT CTATACTAAC CCAGTTCTTG AGAACTTCGA CGGTAGCTTC   840
CGTGGTTCTG CCCAAGGTAT CGAAGGCTCC ATCAGGAGCC CACACTTGAT GGACATCTTG   900
AACAGCATAA CTATCTACAG CGATGCTCAC AGAGGAGAGT ATTACTGGTC TGGACACCAG   960
ATCATGGCCT CTCCAGTTGG ATTCAGCGGG CCCGAGTTTA CCTTTCCTCT CTATGGAACT  1020
ATGGGAAACG CCGCTCCACA ACAACGTATC GTTGCTCAAC TAGGTCAGGG TGTCTACAGA  1080
ACCTTGTCTT CCACCTTGTA CAGAAGACCC TTCAATATCG GTATCAACAA CCAGCAACTT  1140
TCCGTTCTTG ACGGAACAGA GTTCGCCTAT GGAACCTCTT CTAACTTGCC ATCCGCTGTT  1200
TACAGAAAGA GCGGAACCGT TGATTCCTTG GACGAAATCC CACCACAGAA CAACAATGTG  1260
CCACCCAGGC AAGGATTCTC CCACAGGTTG AGCCACGTGT CCATGTTCCG TTCCGGATTC  1320
AGCAACAGTT CCGTGAGCAT CATCAGAGCT CCTATGTTCT CATGGATTCA TCGTAGTGCT  1380
GAGTTCAACA ATATCATTCC TTCCTCTCAA ATCACCCAAA TCCCATTGAC CAAGTCTACT  1440
AACCTTGGAT CTGGAACTTC TGTCGTGAAA GGACCAGGCT TCACAGGAGG TGATATTCTT  1500
```

172

```
AGAAGAACTT CTCCTGGCCA GATTAGCACC CTCAGAGTTA ACATCACTGC ACCACTTTCT    1560

CAAAGATATC GTGTCAGGAT TCGTTACGCA TCTACCACTA ACTTGCAATT CCACACCTCC    1620

ATCGACGGAA GGCCTATCAA TCAGGGTAAC TTCTCCGCAA CCATGTCAAG CGGCAGCAAC    1680

TTGCAATCCG GCAGCTTCAG AACCGTCGGT TTCACTACTC CTTTCAACTT CTCTAACGGA    1740

TCAAGCGTTT TCACCCTTAG CGCTCATGTG TTCAATTCTG GCAATGAAGT GTACATTGAC    1800

CGTATTGAGT TTGTGCCTGC CGAAGTTACC TTCGAGGCTG AGTAC                    1845
```

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page _____159_____ , lines _____21–25_____ .

B. IDENTIFICATION OF DEPOSIT	Further deposits are identified on an additional sheet ☐

Name of depositary institution

Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

1815 North University Street
Peoria, IL 61604
U.S.A.

Date of deposit **4 October 1991 (04.10.91)**	Accession Number **NRRL B 18891**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description	
on page **159**, lines **21-25** .	

B. IDENTIFICATION OF DEPOSIT   Further deposits are identified on an additional sheet ☐

Name of depositary institution

Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

1815 North University Street
Peoria, IL 61604
U.S.A.

Date of deposit	Accession Number
4 October 1991 (04.10.91)	**NRRL B 18892**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page ___159___ , lines ___21-25___ .	

B. IDENTIFICATION OF DEPOSIT   Further deposits are identified on an additional sheet ☐

Name of depositary institution

   Agricultural Research Culture Collection (NRRL)   .

Address of depositary institution *(including postal code and country)*

   1815 North University Street
   Peoria, IL 61604
   U.S.A.

Date of deposit **3 September 1992 (03.09.92)**	Accession Number **NRRL B 18994**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

   We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description

on page ____159____ , lines ____21-25____ .

B. IDENTIFICATION OF DEPOSIT   Further deposits are identified on an additional sheet ☐

Name of depositary institution

Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

**1815 North University Street
Peoria, IL 61604
U.S.A.**

Date of deposit	Accession Number
3 September 1992 (03.09.92)	**NRRL B 18995**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)   177

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page _____ 159 _____ , line s _____ 21-25 _____ .	

B. IDENTIFICATION OF DEPOSIT   Further deposits are identified on an additional sheet ☐

Name of depositary institution

Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

1815 North University Street
Peoria, IL 61604
U.S.A.

Date of deposit **4 September 1992 (04.09.92)**	Accession Number **NRRL B 18996**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

──── For receiving Office use only ────	──── For International Bureau use only ────
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page **159**, lines **21-25**.	

B. IDENTIFICATION OF DEPOSIT   Further deposits are identified on an additional sheet ☐

Name of depositary institution

Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

1815 North University Street
Peoria, IL 61604
U.S.A.

Date of deposit **21 September 1992 (21.09.92)**	Accession Number **NRRL B 18997**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

Form PCT/RO/134 (July 1992)   179

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page __159__ , lines __21-25__ .	

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet ☐

Name of depositary institution

Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

1815 North University Street
Peoria, IL 61604
U.S.A.

Date of deposit	Accession Number
21 September 1992 (21.09.92)	**NRRL B 18998**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on an additional sheet ☐

    We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only	For International Bureau use only
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description
on page   159   , lines   21-25   .

B. IDENTIFICATION OF DEPOSIT   Further deposits are identified on an additional sheet ☐

Name of depositary institution

    Agricultural Research Culture Collection (NRRL)

Address of depositary institution *(including postal code and country)*

    1815 North University Street
    Peoria, IL 61604
    U.S.A.

Date of deposit	Accession Number
21 September 1992 (21.09.92)	**NRRL B 18999**

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)*   This information is continued on an additional sheet ☐

    We request the Expert Solution where available

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leave blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

━━━ For receiving Office use only ━━━	━━━ For International Bureau use only ━━━
☐ This sheet was received with the international application	☐ This sheet was received by the International Bureau on:
Authorized officer	Authorized officer

What is claimed is:

1. A nucleotide sequence comprising a maize optimized coding sequence which encodes a protein capable of killing an insect.

2. A nucleotide sequence of claim 1, wherein said coding sequence encodes a B.t. protein.

3. A nucleotide sequence of claim 2, wherein said coding sequence encodes a CryIA(b) protein.

4. A nucleotide sequence of claim 3, wherein said coding sequence comprises Sequence 3, Sequence 4, or the sequence set forth in Fig. 7.

5. A nucleotide sequence of claim 3, wherein coding sequence encodes a CryIA(b) protein which is heat stable compared to a native CryIA(b) protein.

6. A nucleotide sequence of claim 5, wherein said coding sequence comprises a sequence selected from the group of sequences consisting of Fig. 9, Fig. 11, Fig. 13 and Fig. 15.

7. A nucleotide sequence of claim 2, wherein said coding sequence encodes a CryIB or a CryIA(c) protein.

8. A nucleotide sequence of claim 7, wherein said coding sequence comprises the sequence encoding a CryIB protein set forth in Fig. 6.

9. A nucleotide sequence of claim 1, further comprising a first promoter capable of directing expression of a nucleotide sequence in a plant cell, operably linked to said coding sequence.

10. A nucleotide sequence of claim 9, wherein said

promoter is capable of directing expression of the associated coding sequence in a maize cell.

11.   A nucleotide sequence of claim 9, wherein said promoter is selected from the group consisting of inducible promoters, constitutive promoters, temporal or developmentally-regulated promoters, tissue-preferred, and tissue-specific promoters.

12.   A nucleotide sequence of claim 9, wherein said promoter is selected from the group consisting of a CaMV 35S promoter, CaMV 19S promoter, a PEP carboxylase promoter, a pith-preferred promoter, and a pollen-specific promoter.

13.   A nucleotide sequence of claim 9, wherein said promoter is a pith-preferred promoter comprising the DNA sequence set forth in Fig. 24.

14.   A nucleotide sequence of claim 9, wherein said promoter is a pollen-specific promoter comprising the DNA sequence set forth in Fig. 35.

15.   A nucleotide sequence of claim 9, further comprising a second promoter capable of directing expression of an associated coding sequence in a plant cell, operatively linked to a second coding sequence.

17.   A nucleotide sequence of claim 15, wherein said second promoter is selected from the group consisting of inducible promoters, constitutive promoters, temporal or developmentally-regulated promoters, tissue-preferred, and tissue-specific promoters.

18.   A nucleotide sequence of claim 15, wherein said

second promoter is selected from the group consisting of a CaMV 35S promoter, CaMV 19S promoter, a PEP carboxylase promoter, a pith-preferred promoter, and a pollen-specific promoter.

21. A nucleotide sequence of claim 15, wherein said second coding sequence is a plant optimized coding sequence which encodes a protein capable of killing an insect.

22. A nucleotide sequence of claim 21, wherein said second coding sequence encodes a B.t. protein.

23. A nucleotide sequence of claim 22, wherein said second coding sequence encodes a CryIA(b) protein.

29. A nucleotide sequence of claim 15, wherein said second coding sequence is a marker gene.

30. A recombinant vector, comprising a nucleotide sequence of claim 9.

31. A recombinant vector, comprising a nucleotide sequence of claim 9, wherein said coding sequence encodes a B.t. protein.

32. A recombinant vector of claim 31, wherein the B.t. protein is a CryIA(b) protein.

33. A recombinant vector, comprising a nucleotide sequence of claim 15.

34. A recombinant vector, comprising a nucleotide sequence of claim 18.

35. A recombinant vector, comprising a nucleotide sequence of claim 22.

36. A plant stably transformed with a nucleotide sequence of claim 9.

37.   A plant stably transformed with a nucleotide sequence of claim 12.

38.   A plant of claim 36, wherein said coding sequence encodes a B.t. protein.

39.   A plant of claim 36, wherein said protein is expressed in said plant in an amount sufficient to control Lepidopteran or Coleopteran insects.

40.   A plant of claim 38, wherein the B.t. protein is a CryIA(b) protein.

41.   A plant of claim 38, which expresses the B.t. insecticidal protein in an amount sufficient to control Lepidopteran or Coleopteran pests.

42.   A plant of claim 38, wherein the amount is sufficient to control insects selected from the group consisting of European corn borer, Sugarcane borer, stalk borers, cutworms, armyworms, rootworms, wireworms and aphids.

43.   A plant stably transformed with a nucleotide sequence of claim 15.

44.   A plant stably transformed with a nucleotide sequence of claim 18.

45.   A plant of claim 43, wherein the first and second coding sequences each encode a B.t. protein.

46.   A plant of claim 38, which expresses the B.t. insecticidal proteins in an amount sufficient to control Lepidopteran or Coleopteran pests.

47.   A plant of claim 46, wherein the amount is sufficient to control stalk borers.

185

48. An isolated and purified promoter capable of directing pith-preferred expression of an associated structural gene in a plant.

49. A promoter of claim 48, isolated from a monocot.

50. A promoter of claim 48, isolated from a maize plant.

51. A promoter of claim 48, isolated from a plant tryptophan synthase-alpha (TrpA) subunit gene.

52. A promoter of claim 51, isolated from a maize tryptophan synthase-alpha (TrpA) subunit gene.

53. A promoter of claim 48, comprising the sequence set forth in Figure 24.

54. A recombinant DNA molecule comprising a promoter of claim 48, operably associated with a structural gene encoding a protein of interest.

55. A recombinant DNA molecule of claim 54, wherein said structural gene encodes an insecticidal protein.

56. A recombinant DNA molecule of claim 55, wherein said structural gene encodes a Bacillus thuringiensis protein.

57. A vector, comprising a recombinant DNA molecule of claim 54.

58. A vector of claim 57, wherein said structural gene encodes an insecticidal protein.

59. A vector of claim 57, wherein said structural gene encodes a Bacillus thuringiensis protein.

60. A vector, comprising at least two recombinant DNA molecules of claim 54, wherein at least one of the two

structural genes encodes an insecticidal protein.

61.   A plant stably transformed with recombinant DNA molecule of claim 54.

62.   A plant of claim 61, which is a maize plant.

63.   A purified promoter capable of directing pollen-specific expression of an associated structural gene in a plant, wherein said promoter is isolated from a plant calcium-dependent phosphate kinase (CDPK) gene.

64.   A promoter of claim 63, isolated from a monocot CDPK gene.

65.   A promoter of claim 63, isolated from a maize CDPK gene.

66.   A promoter of claim 65, comprising the sequence set forth in Figure 35.

67.   A recombinant DNA molecule, comprising a promoter of claim 63, operably associated with a structural gene encoding a protein of interest.

68.   A recombinant DNA molecule, of claim 67, wherein said structural gene encodes an insecticidal protein.

69.   A recombinant DNA molecule of claim 68, wherein said structural gene encodes a _Bacillus_ _thuringiensis_ protein.

70.   A vector, comprising at least one recombinant DNA molecule of claim 67.

71.   A vector of claim 70, wherein said structural gene encodes an insecticidal protein.

72.   A vector of claim 71, wherein said structural gene encodes a _Bacillus_ _thuringiensis_ protein.

73.   A vector of claim 70, comprising two recombinant DNA molecules, wherein at least one of the two structural genes encodes an insecticidal protein.

74.   A plant stably transformed with a recombinant DNA molecule of claim 67.

75.   A plant of claim 74, which is maize plant.

76.   A maize plant stably transformed with at least one recombinant DNA molecule;

wherein said DNA molecule comprises a promoter operably linked to a nucleotide sequence encoding an insecticidal protein.

wherein said promoter is capable of directing tissue-preferred or tissue-specific expression of said gene in said maize plant.

77.   A maize plant of claim 76, wherein said gene encodes a <u>Bacillus thuringiensis</u> protein.

78.   A plant of claim 76, wherein said promoter is obtained from a monocot.

79.   A plant of claim 78, wherein said monocot is maize.

80.   A maize plant of claim 76, wherein said promoter is capable of directing pith-preferred expression of said gene in said maize plant.

81.   A plant of claim 76, wherein said promoter is obtained from a plant tryptophan synthase-alpha subunit gene.

82.   A plant of claim 76, wherein said promoter is obtained from a maize tryptophan synthase-alpha subunit gene.

83.   A plant of claim 82, wherein said promoter
comprises the sequence set forth in Figure 24.

84.   A plant of claim 76, wherein said promoter is
capable of directing green tissue-specific expression of said
gene in said maize plant.

85.   A plant of claim 84, wherein said promoter is a
PEP carboxylase promoter.

86.   A maize plant of claim 76, wherein said promoter
is capable of directing pollen-specific expression of said gene
in said maize plant.

87.   A maize plant of claim 76, wherein said promoter
is obtained from a plant calcium-dependent phosphate kinase
gene.

88.   A maize plant of claim 76, wherein said promoter
is obtained from a maize calcium-dependent phosphate kinase
gene.

89.   A maize plant of claim 76, wherein said promoter
comprises the sequence set forth in Figure 35.

90.   A method of producing a maize optimized coding
sequence for an insecticidal B.t. protein, comprising:

determining the amino acid sequence of a predetermined
insecticidal B.t. protein, and

altering the coding sequence of the protein by
substituting codons which are most preferred in maize for
corresponding native codons.

91.   A method of protecting a maize plant against at
least one insect pest, comprising:

stably transforming a maize plant with at least one nucleotide sequence of claim 9, wherein the coding sequence encodes an insecticidal protein; whereby the transformed maize plant expresses the insecticidal protein in an amount sufficient to protect the plant against the pest.

92. The method of claim 91, wherein said insecticidal protein is a B.t. protein.

93. The method of claim 91, wherein said promoter is a tissue-specific or a tissue-preferred promoter.

94. A plant which has been stably transformed with a nucleotide sequence wherein said nucleotide sequence comprises a maize optimized coding sequence for an insecticidal protein, wherein said protein is expressed in said transformed plant at least 100 fold greater than expression of the protein using a native coding sequence.

FIG. 1

```
              10        20        30        40        50        60
               *         *         *         *         *         *
BTHKURHD   ATGGATAACAATCCGAACATCAATGAATGCATTCCTTATAATTGTTTAAGTAACCCTGAA
flsynbt.fin .....C.....C..C........C..G.....C..C..C..C..CC.G..C.....C..G
bssyn      .....C.....C..C........C..G.....C..C..C..C..CC.G..C.....C..G

              70        80        90       100       110       120
               *         *         *         *         *         *
BTHKURHD   GTAGAAGTATTAGGTGGAGAAAGAATAGAAACTGGTTACACCCCAATCGATATTTCCTTG
flsynbt.fin ..G..G..GC.G..C..C..GC.C..C..G..C..C.......C.....C..CAG.C..
bssyn      ..G..G..GC.G..C..C..GC.C..C..G..C..C.......C.....C..CAG.C..

             130       140       150       160       170       180
               *         *         *         *         *         *
BTHKURHD   TCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGGTGCTGGATTTGTGTTAGGACTA
flsynbt.fin AGC..G..C..G..C..GC....C..G..C..G.....C..C..C..C...C.G..C..G
bssyn      AGC..G..C..G..C..GC....C..G..C..G.....C..C..C..C...C.G..C..G

             190       200       210       220       230       240
               *         *         *         *         *         *
BTHKURHD   GTTGATATAATATGGGGAATTTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT
flsynbt.fin ..G..C..C..C.....C..C..C..C...AGC..G........C..C..G..G..G..C
bssyn      ..G..C..C..C.....C..C..C..C...AGC..G........C..C..G..G..G..C

             250       260       270       280       290       300
               *         *         *         *         *         *
BTHKURHD   GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGAACCAAGCCATTTCTAGATTA
flsynbt.fin ..G...C.G..C.....GC.C..C..G..G.....CC.C.....G.....CAGCC.CC.G
bssyn      ..G...C.G..C.....GC.C..C..G..G.....CC.C.....G.....CAGCC.CC.G

             310       320       330       340       350       360
               *         *         *         *         *         *
BTHKURHD   GAAGGACTAAGCAATCTTTATCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAGAT
flsynbt.fin ..G..C..G.....C..G..C.....C.....C..GAGC..CC.C........G..C..C
bssyn      ..G..C..G.....C..G..C.....C.....C..GAGC..CC.C........G..C..C

             370       380       390       400       410       420
               *         *         *         *         *         *
BTHKURHD   CCTACTAATCCAGCATTAAGAGAAGAGATGCGTATTCAATTCAATGACATGAACAGTGCC
flsynbt.fin ..C..C..C..C..CC.GC.C..G........C..C..G.....C...........C...
bssyn      ..C..C..C..C..CC.GC.C..G........C..C..G.....C...........C...

             430       440       450       460       470       480
               *         *         *         *         *         *
BTHKURHD   CTTACAACCGCTATTCCTCTTTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA
flsynbt.fin ..G..C.....C..C..C..G..C..C..G..G..C..C..G..G..C..GC.GAGC..G
bssyn      ..G..C.....C..C..C..G..C..C..G..G..C..C..G..G..C..GC.GAGC..G

             490       500       510       520       530       540
               *         *         *         *         *         *
BTHKURHD   TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAGATGTTTCAGTGTTTGGACAA
flsynbt.fin ..C..G..G..C..C..CC.G..CC.GAGC..GC..C.C..C..CAGC.....C..C..G
bssyn      ..C..G..G..C..C..CC.G..CC.GAGC..GC..C.C..C..CAGC.....C..C..G

             550       560       570       580       590       600
               *         *         *         *         *         *
BTHKURHD   AGGTGGGGATTTGATGCCGCGACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT
flsynbt.fin C.C.....C..C..C.....C...C.....C..C..C..C..C..CC.G..CC.C..G..C
bssyn      C.C.....C..C..C.....C...C.....C..C..C..C..C..CC.G..CC.C..G..C
```

1/94

FIG. 1 (CONT)

```
              121.      1220      1230      124c      1250      1260
                *         *         *         *         *         *
BTHKURHD    TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATACCGCCACAGAATAACAACGTG
flsynbt.fin ...C.C..G.....C..C..G..CAGC.....C..G..C..C..T.....C.........
bssyn       ...C.C..G.....C..C..G..CAGC.....C..G..C..C..T.....C.........
 ‘
 ;
              1270      1280      1290      1300      1310      1320
                *         *         *         *         *         *
BTHKURHD    CCACCTAGGCAAGGATTTAGTCATCGATTAAGCCATGTTTCAATGTTTCGTTCAGGCTTT
flsynbt.fin ......C.A..G..C..C..C..C..TC.G.....C..GAGC.....C..CAGT.....C
bssyn  −    ......C.A..G..C..C..C..C..TC.G.....C..GAGC.....C..CAGT.....C

              1330      1340      1350      1360      1370      1380
                *         *         *         *         *         *
BTHKURHD    AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCTCTTGGATACATCGTAGTGCT
flsynbt.fin ..C..C..C..C..G..C..C..CC.T..A.........AGC.....T..C..C.....C
bssyn       ..C..C..C..C..G..C..C..CC.T..A.........AGC.....T..C..C.....C

              1390      1400      1410      1420      1430
                *         *         *         *         *
BTHKURHD    GAATTTAATAATATAATTCCTTCATCA--CAAATTACACAAATACCTTTAACAAAATCTA
flsynbt.fin ..G..C..C..C..C..C..C..--..G..GC..G..C..C..G..C..CC.G..C..GAGC.
bssyn       ..G..C..C..C..C..C..C..--..G..GC..G..C..C..G..C..CC.G..C..GAGC.

            :440      1450      1460      147c      1480      :49 c
               *         *         *         *         *         *
BTHKURHD    CTAATCTTGGCTCTGGAACTTCTGTCGTTAAAGGACCAGGATTTACAGGAGGAGATATTC
flsynbt.fin .C..C..G...AGC..C..CAGC..G..G..G..C..C..C..C..C..C..C..C..C.
bssyn       .C..C..G...AGC..C..CAGC..G..G..G..C..C..C..C..C..C..C..C..C.

            1500      1510      1520      1530      1540      1550
               *         *         *         *         *         *
BTHKURHD    TTCGAAGAACTTCACCTGGCCAGATTTCAACCTTAAGAGTAAATATTACTGCACCATTAT
flsynbt.fin .G..CC.C..CAGC..C........CAGC...C.GC.C..G..C..C..C..C..CC.GA
bssyn       .G..CC.C..CAGC..C........CAGC...C.GC.C..G..C..C..C..C..CC.GA

            :560      1570      1580      1590      1600      :610
               *         *         *         *         *         *
BTHKURHD    CACAAAGATATCGGGTAAGAATTCGCTACGCTTCTACCACAAATTTACAATTCCATACAT
flsynbt.fin GC..GC.C..C..C..CC.C..C........CAGC.....C..CC.G..G.....C..CA
bssyn       GC..GC.C..C..C..CC.C..C........CAGC.....C..CC.G..G.....C..CA

            1620      1630      1640      1650      1660      1670
               *         *         *         *         *         *
BTHKURHD    CAATTGACGGAAGACCTATTAATCAGGGGAATTTTTCAGCAACTATGAGTAGTGGGAGTA
flsynbt.fin GC..C.....CC.C..C..C..C..:..C..C..CAGC..C..C.....C..C..C..C.
bssyn       GC..C.....CC.C..C..C..C..C..C..C..CAGC..C..C.....C..C..C..C.

            1680      1690      1700      1710      1720      1730
               *         *         *         *         *         *
BTHKURHD    ATTTACAGTCCGGAAGCTTTAGGACTGTAGGTTTTACTACTCCGTTTAACTTTTCAAATG
flsynbt.fin .CC.G...AG...C.....CC.C..C..G..C..C..C..C..C..C.....CAGC..C.
bssyn       .CC.G...AG...C.....CC.C..C..G..C..C..C..C..C..C.....CAGC..C.

            1740      1750      1760      1770      1780      1790 ·
               *         *         *         *         *         *
BTHKURHD    GATCAAGTGTATTTACGTTAAGTGCTCATGTCTTCAATTCAGGCAATGAAGTTTATATAG
flsynbt.fin .CAGC..C..G..C..CC.G..C..C..C..G.....CAGC.....C..G..G..C..C.
bssyn       .CAGC..C..G..C..CC.G..C..C..C..G.....CAGC.....C..G..G..C..C.
```

2/94

FIG. 1 (CONT)

```
          1800      310       1820      1830      ⌐40       1850
            *         *         *         *         *         *
BTHKURHD  ATCGAATTGAATTTGTTCCGGCAGAAGTAACCTTTGAGGCAGAATATGATTTAGAAAGAG
flsynbt.fin  .C..C..C..G..C..G..C..C..G..G.....C.....C..G..C..CC.G..G..G.
bssyr.       .C..C..C..G..C..G..C..C..G..G.....C.....C..G..C..CC.G..G..G.

          1860      1870      1880      1890      1900      1910
            *         *         *         *         *         *
BTHKURHD  CACAAAAGGCGGTGAATGAGCTGTTTACTTCTTCCAATCAAATCGGGTTAAAAACAGATG
flsynbt.fin  .T..G.....C.....C........C..CAGCAG...C..G.....CC.G..G..C..C.
bssyn        .T..G.....C.....C........C..CAGCAG...C..G.....CC.G..G..C..C.

          1920      1930      1940      1950      1960      1970
            *         *         *         *         *         *
BTHKURHD  TGACGGATTATCATATTGATCAAGTATCCAATTTAGTTGAGTGTTTATCTGATGAATTTT
flsynbt.fin  ....C..C..C..C..C.....G..GAG...CC.G..G.....CC.GAGC..C..G..C.
bssyn        ....C..C..C..C..C....-----------------------------------------

          1980      1990      2000      2010      2020      2030
            *         *         *         *         *         *
BTHKURHD  GTCTGGATGAAAAAAAAGAATTGTCCGAGAAAGTCAAACATGCGAAGCGACTTAGTGATG
flsynbt.fin  .C.....C..G..G..G..GC..AG......G..G..G..C..C.....C..G..C..C.
bssyn        -------------------------------------------------------------

          2040      2050      2060      2070      2080      2090
            *         *         *         *         *         *
BTHKURHD  AGCGGAATTTACTTCAAGATCCAAACTTTAGAGGGATCAATAGACAACTAGACCGTGGCT
flsynbt.fin  ....C..CC.G..G..G..C..C.....CC.C..C.....CC.C..G..G.....C....
bssyn        -------------------------------------------------------------

          2100      2110      2120      2130      2140      2150
            *         *         *         *         *         *
BTHKURHD  GGAGAGGAAGTACGGATATTACCATCCAAGGAGGCGATGACGTATTCAAAGAGAATTACG
flsynbt.fin  ..C.C..C..C..C..C..C.........G..C.....C.....G.....G.....C....
bssyn        -------------------------------------------------------------

          2160      2170      2180      2190      2200      2210
            *         *         *         *         *         *
BTHKURHD  TTACGCTATTGGGTACCTTTGATGAGTGCTATCCAACGTATTTATATCAAAAAATAGATG
flsynbt.fin  .G..C..GC....C.....C..C.........C..C..C..CC.G..C..G..G..C..C.
bssyn        -------------------------------------------------------------

          2220      2230      2240      2250      2260      2270
            *         *         *         *         *         *
BTHKURHD  AGTCGAAATTAAAAGCCTATACCCGTTACCAATTAAGAGGGTATATCGAAGATAGTCAAG
flsynbt.fin  ..AGC..GC.G..G.....C.....·C.....GC.GC.C..C..C.....G..C..C..G.
bssyn        -------------------------------------------------------------

          2280      2290      2300      2310      2320      2330
            *         *         *         *         *         *
BTHKURHD  ACTTAGAAATCTATTTAATTCGCTACAATGCCAAACACGAAACAGTAAATGTGCCAGGTA
flsynbt.fin  ..C.G..G.....CC.G..C........C.....G.....G..C..G..C.....C..C.
bssyn        -------------------------------------------------------------

          2340      2350      2360      2370      2380      2390 ·
            *         *         *         *         *         *
BTHKURHD  CGGGTTCCTTATGGCCGCTTTCAGCCCCAAGTCCAATCGGAAAATGTGCCCATCATTCCC
flsynbt.fin  .C..CAG.C.G.....C...GAGC.....C..C..C.....C..G..C.....C...CAG..
bssyn        -------------------------------------------------------------
```

3/94

FIG. 1 (CONT)

```
        2400       2410       2420       2430       2440       2450
          *          *          *          *          *          *
BTHKURHD  ATCATTTCTCCTTGGACATTGATGTTGGATGTACAGACTTAAATGAGGACTTAGGTGTAT
flsynbt.fin  .C..C...AG.C.......C..C..G..C..C..C...C.G..C......C.G..C..G.
bssyn     ------------------------------------------------------------

        2460       2470       2480       2490       2500       2510
          *          *          *          *          *          *
BTHKURHD  GGGTGATATTCAAGATTAAGACGCAAGATGGCCATGCAAGACTAGGAAATCTAGAATTTC
flsynbt.fin  .......C........C.....C..G..C......C..CC.C..G..C..C..G..G..C.
bssyn     ------------------------------------------------------------

        2520       2530       2540       2550       2560       2570
          *          *          *          *          *          *
BTHKURHD  TCGAAGAGAAACCATTAGTAGGAGAAGCACTAGCTCGTGTGAAAAGAGCGGAGAAAAAAT
flsynbt.fin  .G..G.....G..CC.G..G..C..G..C..G..C..C.....GC.C..C.....G..G.
bssyn     ------------------------------------------------------------

        2580       2590       2600       2610       2620       2630
          *          *          *          *          *          *
BTHKURHD  GGAGAGACAAACGTGAAAAATTGGAATGGGAAACAAATATTGTTTATAAAGAGGCAAAAG
flsynbt.fin  ..C.C.....G..C..G..GC....G.....G..C..C..C.G..C..G.....C..G.
bssyn     ------------------------------------------------------------

        2640       2650       2660       2670       2680       2690
          *          *          *          *          *          *
BTHKURHD  AATCTGTAGATGCTTTATTTGTAAACTCTCAATATGATAGATTACAAGCGGATACCAACA
flsynbt.fin  .GAGC..G..C..CC.G..C..G...AGC..G..C..CC.CC.G..G..C..C.......
bssyn     ------------------------------------------------------------

        2700       2710       2720       2730       2740       2750
          *          *          *          *          *          *
BTHKURHD  TCGCGATGATTCATGCGGCAGATAAACGCGTTCATAGCATTCGAGAAGCTTATCTGCCTG
flsynbt.fin  ....C.....C..C..C..C..C..G.....G..C.........C..G..C..C.....C.
bssyn     ------------------------------------------------------------

        2760       2770       2780       2790       2800       2810
          *          *          *          *          *          *
BTHKURHD  AGCTGTCTGTGATTCCGGGTGTCAATGCGGCTATTTTTGAAGAATTAGAAGGGCGTATTT
flsynbt.fin  .....AGC.....C..C..C..G..C..C..C..C..C..G..GC.G..G..C..C..C.
bssyn     ------------------------------------------------------------

        2820       2830       2840       2850       2860       2870
          *          *          *          *          *          *
BTHKURHD  TCACTGCATTCTCCCTATATGATGCGAGAAATGTCATTAAAAATGGTGATTTTAATAATG
flsynbt.fin  ....C..C...AG...G..C..C..CC.C..C..G..C..G..C..C..C..C..C..C.
bssyn     ------------------------------------------------------------

        2880       2890       2900       2910       2920       2930
          *          *          *          *          *          *
BTHKURHD  GCTTATCCTGCTGGAACGTGAAAGGGCATGTAGATGTAGAAGAACAAAACAACCACCGTT
flsynbt.fin  ..C.GAG................G..C..C..G..C..G..G..G..G...........CA
bssyn     ------------------------------------------------------------

        2940       2950       2960       2970       2980       2990
          *          *          *          *          *          *
BTHKURHD  CGGTCCTTGTTGTTCCGGAATGGGAAGCAGAAGTGTCACAAGAAGTTCGTGTCTGTCCGG
flsynbt.fin  GC..G..G..G..G..C..G.....G..C..G...AGC..G..G..G..C..G..C..C.
bssyn     ------------------------------------------------------------
```

FIG. 1 (CONT)

```
                3000      3010      3020      3030      3040      3050
                  *         *         *         *         *         *
BTHKURHD    GTCGTGGCTATATCCTTCGTGTCACAGCGTACAAGGAGGGATATGGAGAAGGTTGCGTAA
flsynbt.fin .C..C.....C.....G..C..G..C..C...........C..C..C..G..C.....G.
bssyn       ------------------------------------------------------------


                3060      3070      3080      3090      3100      3110
                  *         *         *         *         *         *
BTHKURHD    CCATTCATGAGATCGAGAACAATACAGACGAACTGAAGTTTAGCAACTGTGTAGAAGAGG
flsynbt.fin ....C..C...............C..C.....G..C.....C........C..G..G....
bssyn       ------------------------------------------------------------


                3120      3130      3140      3150      3160      3170
                  *         *         *         *         *         *
BTHKURHD    AAGTATATCCAAACAACACGGTAACGTGTAATGATTATACTGCGACTCAAGAAGAATATG
flsynbt.fin .G..G..C..C.........C..G..C..C..C..C..C..C..C..C..G..G..G..C.
bssyn       ------------------------------------------------------------


                3180      3190      3200      3210      3220      3230
                  *         *         *         *         *         *
BTHKURHD    AGGGTACGTACACTTCTCGTAATCGAGGATATGACGGAGCCTATGAAAGCAATTCTTCTG
flsynbt.fin ....C..C.....CAGC..C..C..C..C..C.....C.....C..G.....CAGCAGC.
bssyn       ------------------------------------------------------------


                3240      3250      3260      3270      3280      3290
                  *         *         *         *         *         *
BTHKURHD    TACCAGCTGATTATGCATCAGCCTATGAAGAAAAAGCATATACAGATGGACGAAGAGACA
flsynbt.fin .G..C..C..C..C..CAGC.....C..G..G..G..C..C..C..C..C..CC.C....
bssyn       ------------------------------------------------------------


                3300      3310      3320      3330      3340      3350
                  *         *         *         *         *         *
BTHKURHD    ATCCTTGTGAATCTAACAGAGGATATGGGGATTACACACCACTACCAGCTGGCTATGTGA
flsynbt.fin .C..C..C..GAGC...C.C..C..C..C..C.....C..C..G..C..C.....C....
bssyn       ------------------------------------------------------------


                3360      3370      3380      3390      3400      3410
                  *         *         *         *         *         *
BTHKURHD    CAAAAGAATTAGAGTACTTCCCAGAAACCGATAAGGTATGGATTGAGATCGGAGAAACGG
flsynbt.fin .C..G..GC.G...........C..G.....C.....G.....C........C..G..C.
bssyn       ------------------------------------------------------------


                3420      3430      3440      3450      3460
                  *         *         *         *         *
BTHKURHD    AAGGAACATTCATCGTGGACAGCGTGGAATTACTTCTTATGGAGGAATAA
flsynbt.fin .G..C..C...................GC.G..G..G........G..G
bssyn       ------------------------------------------...TG..G
```

FIG. 2

```
          10         .20        30         4L         50         60
           *          *          *          *          *          *
BTHKURHD  ATGGATAACAATCCGAACATCAATGAATGCATTCCTTATAATTGTTTAAGTAACCCTGAA
bssyn     .....C.....C..C.........C..G.....C..C..C..C..CC.G..C.....C..G

          70         80         90         100        110        120
           *          *          *          *          *          *
BTHKURHD  GTAGAAGTATTAGGTGGAGAAAGAATAGAAACTGGTTACACCCCAATCGATATTTCCTTG
bssyn     ..G..G..GC.G..C...C..GC.C..C..G..C..C.........C.....C..CAG.C..

          130        140        150        160        170        180
           *          *          *          *          *          *
BTHKURHD  TCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGGTGCTGGATTTGTGTTAGGACTA
bssyn     AGC..G..C..G..C..GC....C..G..C..G.....C..C..C..C...C.G..C..G

          190        200        210        220        230        240
           *          *          *          *          *          *
BTHKURHD  GTTGATATAATATGGGGAATTTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT
bssyn     ..G..C..C..C.....C..C..C..C...AGC..G.........C..C..G..G..G..C

          250        260        270        280        290        300
           *          *          *          *          *          *
BTHKURHD  GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGAACCAAGCCATTTCTAGATTA
bssyn     ..G...C.G..C.....GC.C..C..G..G.....CC.C.....G.....CAGCC.CC.G

          310        320        330        340        350        360
           *          *          *          *          *          *
BTHKURHD  GAAGGACTAAGCAATCTTTATCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAGAT
bssyn     ..G..C..G.....C..G..C.....C.....C..GAGC..CC.C........G..C..C

          370        380        390        400        410        420
           *          *          *          *          *          *
BTHKURHD  CCTACTAATCCAGCATTAAGAGAAGAGATGCGTATTCAATTCAATGACATGAACAGTGCC
bssyn     ..C..C..C..C..CC.GC.C..G........C..C..G.....C...........C...

          430        440        450        460        470        480
           *          *          *          *          *          *
BTHKURHD  CTTACAACCGCTATTCCTCTTTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA
bssyn     ..G..C.....C..C..C..G..C..C..G..G..C..C..G..G..C..GC.GAGC..G

          490        500        510        520        530        540
           *          *          *          *          *          *
BTHKURHD  TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAGATGTTTCAGTGTTTGGACAA
bssyn     ..C..G..G..C..C..CC.G..CC.GAGC..GC..C.C..C..CAGC.....C..C..G

          550        560         570        580        590        600
           *          *          *          *          *          *
BTHKURHD  AGGTGGGGATTTGATGCCGCGACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT
bssyn     C.C.....C..C..C.....C..C.....C..C..C..C..C..CC.G..CC.C..G..C

          610        620        630        640        650        660
           *          *          *          *          *          *
BTHKURHD  GGCAACTATACAGATCATGCTGTACGCTGGTACAATACGGGATTAGAGCGTGTATGGGGA
bssyn     ........C..C..C..C..C..G...........C..C..CC.G.....C..G.....T

          670        680        690        700        710        720
           *          *          *          *          *          *
BTHKURHD  CCGGATTCTAGAGATTGGATAAGATATAATCAATTTAGAAGAGAATTAACACTAACTGTA
bssyn     ..C..CAGCC.C..C.....C..G..C..C..G..CC.CC.C..GC.G..C..G..C..G
```

FIG. 2 (CONT)

```
            610       620       630       64u       650       660
             *         *         *         *         *         *
BTHKURHD    GGCAACTATACAGATCATGCTGTACGCTGGTACAATACGGGATTAGAGCGTGTATGGGGA
flsynbt.fin ........C..C..C..C..C..G............C..C..CC.G.....C..G.....T
bssyn       ........C..C..C..C..C..G............C..C..CC.G.....C..G.....T

            670       680       690       700       710       720
             *         *         *         *         *         *
BTHKURHD    CCGGATTCTAGAGATTGGATAAGATATAATCAATTTAGAAGAGAATTAACACTAACTGTA
flsynbt.fin ..C..CAGCC.C..C.....C..G..C..C..G..CC.CC.C..GC.G..C..G..C..G
bssyn       ..C..CAGCC.C..C.....C..G..C..C..G..CC.CC.C..GC.G..C..G..C..G

            730       740       750       760       770       780
             *         *         *         *         *         *
BTHKURHD    TTAGATATCGTTTCTCTATTTCCGAACTATGATAGTAGAACGTATCCAATTCGAACAGTT
flsynbt.fin C.G..C.....GAGC..G..C..C.....C..C..CC.C..C..C..C..C..C..C..G
bssyn       C.G..C.....GAGC..G..C..C.....C..C..CC.C..C..C..C..C..C..C..G

            790       800       810       820       830       840
             *         *         *         *         *         *
BTHKURHD    TCCCAATTAACAAGAGAAATTTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT
flsynbt.fin AG...GC.G..CC.C..G.....C..C.....C..GC.G..G..C..C..C..C..C..C
bssyn       AG...GC.G..CC.C..G.....C..C.....C..GC.G..G..C..C..C..C..C..C

            850       860       870       880       890       900
             *         *         *         *         *         *
BTHKURHD    CGAGGCTCGGCTCAGGGCATAGAAGGAAGTATTAGGAGTCCACATTTGATGGATATACTT
flsynbt.fin ..C...AGC..C........C..G..C..C..CC.C..C..C..CC.......C..C..G
bssyn       ..C...AGC..C........C..G..C..C..CC.C..C..C..CC.......C..C..G

            910       920       930       940       950       960
             *         *         *         *         *         *
BTHKURHD    AACAGTATAACCATCTATACGGATGCTCATAGAGGAGAATATTATTGGTCAGGGCATCAA
flsynbt.fin .....C..C........C..C..C..C..CC.C..C..G..C..C...AGC..C..C..G
bssyn       .....C..C........C..C..C..C..CC.C..C..G..C..C...AGC..C..C..G

            970       980       990      1000      1010      1020
             *         *         *         *         *         *
BTHKURHD    ATAATGGCTTCTCCTGTAGGGTTTTCGGGGCCAGAATTCACTTTTCCGCTATATGGAACT
flsynbt.fin ..C.....CAGC..C..C..C..CAGC..C..C..G.....C..C..C..G..C..C..C
bssyn       ..C.....CAGC..C..C..C..CAGC..C..C..G.....C..C..C..G..C..C..C

           1030      1040      1050      1060      1070      1080
             *         *         *         *         *         *
BTHKURHD    ATGGGAAATGCAGCTCCACAACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA
flsynbt.fin .....C..C..T..A..T..G..G..C..C..G..A..G..G..C.....A.....CC.C
bssyn       .....C..C..T..A..T..G..G..C..C..G..A..G..G..C.....A.....CC.C

           1090      1100      1110      1120      1130      1140
             *         *         *         *         *         *
BTHKURHD    ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAGGGATAAATAATCAACAACTA
flsynbt.fin ..CC.GAGCAG...CC.G..CC.TC.......C..C..C..C..C..C..C..G..G..G
bssyn       ..CC.GAGCAG...CC.G..CC.TC.......C..C..C..C..C..C..C..G..G..G

           1150      1160      1170      1180      1190      1200
             *         *         *         *         *         *
BTHKURHD    TCTGTTCTTGACGGGACAGAATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTA
flsynbt.fin AGC..G..G.....C..C..G..C..C..C..C...AG.AGC..CC....CAG...C..G
bssyn       AGC..G..G.....C..C..G..C..C..C..C...AG.AGC..CC....CAG...C..G
```

FIG. 2 (CONT)

```
        7..      740        750       76        770       780
         *        *          *         *         *         *
BTHKURHD  TTAGATATCGTTTCTCTATTTCCGAACTATGATAGTAGAACGTATCCAATTCGAACAGTT
bssyn     C.G..C.....GAGC..G..C..C.....C..C..CC.C...C..C..C..C..C..G

        790       800       810       820       830       840
         *         *         *         *         *         *
BTHKURHD  TCCCAATTAACAAGAGAAATTTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT
bssyn     AG...GC.G..CC.C..G.....C..C.....C..GC.G..G..C..C..C..C..C

        850       860       870       880       890       900
         *         *         *         *         *         *
BTHKURHD  CGAGGCTCGGCTCAGGGCATAGAAGGAAGTATTAGGAGTCCACATTTGATGGATATACTT
bssyn     ..C...AGC..C.........C..G..C..C..CC.C..C..C..CC.......C..C..G

        910       920       930       940       950       960
         *         *         *         *         *         *
BTHKURHD  AACAGTATAACCATCTATACGGATGCTCATAGAGGAGAATATTATTGGTCAGGGCATCAA
bssyn     .....C..C.........C..C..C..C..CC.C..C..G..C..C...AGC..C..C..G

        970       980       990       1000      1010      1020
         *         *         *         *         *         *
BTHKURHD  ATAATGGCTTCTCCTGTAGGGTTTTCGGGGCCAGAATTCACTTTTCCGCTATATGGAACT
bssyn     ..C.....CAGC..C..C..C..CAGC..C..C..G.....C..C..C..G..C..C..C

        1030      1040      1050      1060      1070      1080
         *         *         *         *         *         *
BTHKURHD  ATGGGAAATGCAGCTCCACAACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA
bssyn     .....C..C..T..A..T..G..G..C..C..G..A..G..G..C.....A.....CC.C

        1090      1100      1110      1120      1130      1140
         *         *         *         *         *         *
BTHKURHD  ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAGGGATAAATAATCAACAACTA
bssyn     ..CC.GAGCAG...CC.G..CC.TC.......C..C..C..C..C..C..C..G..G..G

        1150      1160      1170      1180      1190      1200
         *         *         *         *         *         *
BTHKURHD  TCTGTTCTTGACGGGACAGAATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTA
bssyn     AGC..G..G.....C..C..G..C..C..C..C...AG.AGC..CC....CAG...C..G

        1210      1220      1230      1240      1250      1260
         *         *         *         *         *         *
BTHKURHD  TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATACCGCCACAGAATAACAACGTG
bssyn     ...C.C..G.....C..C..G..CAGC.....C..G..C..C..T.....C.........

        1270      1280      1290      1300      1310      1320
         *         *         *         *         *         *
BTHKURHD  CCACCTAGGCAAGGATTTAGTCATCGATTAAGCCATGTTTCAATGTTTCGTTCAGGCTTT
bssyn     ......C.A..G..C..C..C..C..TC.G.....C..GAGC.....C..CAGT.....C

        1330      1340      1350      1360      1370      1380
         *         *         *         *         *         *
BTHKURHD  AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCTCTTGGATACATCGTAGTGCT
bssyn     ..C..C..C..C..G..C..C..CC.T..A.........AGC.....T..C..C.....C

        1390      1400      1410      1420      1430
         *         *         *         *         *
BTHKURHD  GAATTTAATAATATAATTCCTTCATCA--CAAATTACACAAATACCTTTAACAAAATCTA
bssyn     ..G..C..C..C..C..C..--..G...GC.G..C..C..G..C..CC.G..C..GAGC.
```

FIG. 2 (CONT)

```
                1440      1450      1460      1470      1480      1490
                  *         *         *         *         *         *
BTHKURHD        CTAATCTTGGCTCTGGAACTTCTGTCGTTAAAGGACCAGGATTTACAGGAGGAGATATTC
bssyn           .C..C..G...AGC..C..CAGC..G..G..G..C..C..C..C..C..C..C..C..C.


                1500      1510      1520      1530      1540      1550
                  *         *         *         *         *         *
BTHKURHD        TTCGAAGAACTTCACCTGGCCAGATTTCAACCTTAAGAGTAAATATTACTGCACCATTAT
bssyn           .G..CC.C..CAGC..C........CAGC...C.GC.C..G..C..C..C..C..CC.GA


                1560      1570      1580      1590      1600      1610
                  *         *         *         *         *         *
BTHKURHD        CACAAAGATATCGGGTAAGAATTCGCTACGCTTCTACCACAAATTTACAATTCCATACAT
bssyn           GC..GC.C..C..C..CC.C..C........CAGC.....C..CC.G..G.....C..CA


                1620      1630      1640      1650      1660      1670
                  *         *         *         *         *         *
BTHKURHD        CAATTGACGGAAGACCTATTAATCAGGGGAATTTTTCAGCAACTATGAGTAGTGGGAGTA
bssyn           GC..C.....CC.C..C..C..C.....C..C..CAGC..C..C.....C..C..C..C.


                1680      1690      1700      1710      1720      1730
                  *         *         *         *         *         *
BTHKURHD        ATTTACAGTCCGGAAGCTTTAGGACTGTAGGTTTTACTACTCCGTTTAACTTTTCAAATG
bssyn           .CC.G...AG...C.....CC.C..C..G..C..C..C..C..C.....CAGC..C.


                1740      1750      1760      1770      1780      1790
                  *         *         *         *         *         *
BTHKURHD        GATCAAGTGTATTTACGTTAAGTGCTCATGTCTTCAATTCAGGCAATGAAGTTTATATAG
bssyn           .CAGC..C..G..C..CC.G..C..C..C..G.....CAGC.....C..G..G..C..C.


                1800      1810      1820      1830      1840      1850
                  *         *         *         *         *         *
BTHKURHD        ATCGAATTGAATTTGTTCCGGCAGAAGTAACCTTTGAGGCAGAATATGATTTAGAAAGAG
bssyn           .C..C..C..G..C..G..C..C..G..G.....C.....C..G..C..CC.G..G..G.


                1860      1870      1880      1890      1900      1910
                  *         *         *         *         *         *
BTHKURHD        CACAAAAGGCGGTGAATGAGCTGTTTACTTCTTCCAATCAAATCGGGTTAAAAAACAGATG
bssyn           .T..G.....C.....C........C..CAGCAG...C..G.....CC.G..G..C..C.


                1920      1930      1940
                  *         *         *
BTHKURHD        TGACGGATTATCATATTGATCAAGTATCC
bssyn           ....C..C..C..C..C.....G..G.AG
```

FIG. 3

```
                 10        20        30        40        50        60
                  *         *         *         *         *         *
syn1T.mze    ATGGACAACAACCCCAACATCAACGAGTGCATCCCCTACAACTGCCTGAGCAACCCCGAG
bssyn        ............................................................
synful.mod   ............................................................

                 70        80        90       100       110       120
                  *         *         *         *         *         *
syn1T.mze    GTGGAGGTGCTGGGCGGCGAGCGCATCGAGACCGGCTACACCCCCATCGACATCAGCCTG
bssyn        ............................................................
synful.mod   ............................................................

                130       140       150       160       170       180
                  *         *         *         *         *         *
syn1T.mze    AGCCTGACCCAGTTCCTGCTGAGCGAGTTCGTGCCCGGCGCCGGCTTCGTGCTGGGCCTG
bssyn        ............................................................
synful.mod   ............................................................

                190       200       210       220       230       240
                  *         *         *         *         *         *
syn1T.mze    GTGGACATCATCTGGGGCATCTTCGGCCCCCAGCCAGTGGGACGCCTTCCTGGTGCAGATC
bssyn        ............................................................
synful.mod   ............................................................

                250       260       270       280       290       300
                  *         *         *         *         *         *
syn1T.mze    GAGCAGCTGATCAACCAGCGCATCGAGGAGTTCGCCCGCAACCAGGCCATCAGCCGCCTG
bssyn        ............................................................
synful.mod   ............................................................

                310       320       330       340       350       360
                  *         *         *         *         *         *
syn1T.mze    GAGGGCCTGAGCAACCTGTACCAGATCTACGCCGAGAGCTTCCGCGAGTGGGAGGCCGAC
bssyn        ......................A.....................................
synful.mod   ......................A.....................................

                370       380       390       400       410       420
                  *         *         *         *         *         *
syn1T.mze    CCCACCAACCCCGCCCTGCGCGAGGAGATGCGCATCCAGTTCAACGACATGAACAGCGCC
bssyn        ............................................................
synful.mod   ............................................................

                430       440       450       460       470       480
                  *         *         *         *         *         *
syn1T.mze    CTGACCACCGCCATCCCCCTGTTCGCCGTGCAGAACTACCAGGTGCCCCTGCTGAGCGTG
bssyn        ............................................................
synful.mod   ............................................................

                490       500       510       520       530       540
                  *         *         *         *         *         *
syn1T.mze    TACGTGCAGGCCGCCAACCTGCACCTGAGCGTGCTGCGCGACGTGAGCGTGTTCGGCCAG
bssyn        ..................................................C.........
synful.mod   ..................................................C.........

                550       560       570       580       590       600
                  *         *         *         *         *         *
syn1T.mze    CGCTGGGGGCTTCGACGCCGCCACCATCAACAGCCGCTACAACGACCTGACCCGCCTGATC
bssyn        ............................................................
synful.mod   ............................................................
```

FIG. 3 (CONT)

```
         6.0        620        630        64J        650        660
          .          *          *          *          *          *
syn1T.mze GGCAACTACACCGACCACGCCGTGCGCTGGTACAACACCGGCCTGGAGCGCGTGTGGGGC
bssyn     ...........................................................T
synful.mod ..........................................................T

         670        680        690        700        710        720
          *          *          *          *          *          *
syn1T.mze CCCGACAGCCGCGACTGGATCCGCTACAACCAGTTCCGCCGCGAGCTGACCCTGACCGTG
bssyn     ......................A.G..................................
synful.mod ......................A.G..................................

         730        740        750        760        770        780
          *          *          *          *          *          *
syn1T.mze CTGGACATCGTGAGCCTGTTCCCCAACTACGACAGCCGCACCTACCCCATCCGCACCGTG
bssyn     ............................................................
synful.mod ............................................................

         790        800        810        820        830        840
          *          *          *          *          *          *
syn1T.mze AGCCAGCTGACCCGCGAGATCTACACCAACCCCGTGCTGGAGAACTTCGACGGCAGCTTC
bssyn     ...........................T................................
synful.mod ..........................T.................................

         850        860        870        880        890        900
          *          *          *          *          *          *
syn1T.mze CGCGGCAGCGCCCAGGGCATCGAGGGCAGCATCCGCAGCCCCCACCTGATGGACATCCTG
bssyn     ............................................................
synful.mod ............................................................

         910        920        930        940        950        960
          *          *          *          *          *          *
syniT.mze AACAGCATCACCATCTACACCGACGCCCACCGCGGCGAGTACTACTGGAGCGGCCACCAG
bssyn     ............................................................
synful.mod ............................................................

         970        980        990        1000       1010       1020
          *          *          *          *          *          *
syn1T.mze ATCATGGCCAGCCCCGTGGGCTTCAGCGGCCCCGAGTTCACCTTCCCCCTGTACGGCACC
bssyn     ......................C.....................................
synful.mod ......................C.....................................

         1030       1040       1050       1060       1070       1080
          *          *          *          *          *          *
syn1T.mze ATGGGCAACGCCGCCCCCCCAGCAGCGCATCGTGGCCCAGCTGGGCCAGGGCGTGTACCGC
bssyn     ...........T..A..T....................A.................A.........
synful.mod ...........T..A..T....................A.................A.........

         1090       1100       1110       1120       1130       1140
          *          *          *          *          *          *
syn1T.mze ACCCTGAGCAGCACCCTGTACCGCCGCCCCCTTCAACATCGGCATCAACAACCAGCAGCTG
bssyn     ............................T..A..T..........................
synful.mod ............................T..A..T..........................

         1150       1160       1170       1180       1190       1200
          *          *          *          *          *          *
syn1T.mze AGCGTGCTGGACGGCACCGAGTTCGCCTACGGCACCAGCAGCAACCTGCCCAGCGCCGTG
bssyn     ............................................................
synful.mod ............................................................
```

11/94

FIG. 3 (CONT)

```
            1210      1220      1230      1240      1250      1260
              *         *         *         *         *         *
synlT.mze  TACCGCAAGAGCGGCACCGTGGACAGCCTGGACGAGATCCCCCCCCCAGAACAACAACGTG
bssyn      .............................................T..............
synful.mod .............................................T..............

            1270      1280      1290      1300      1310      1320
              *         *         *         *         *         *
synlT.mze  CCCCCCCCGCCAGGGCTTCAGCCACCGCCTGAGCCACGTGAGCATGTTCCGCAGCGGCTTC
bssyn      ..A..T..A..................T.......................T......
synful.mod ..A..T..A..................T.......................T......

            1330      1340      1350      1360      1370      1380
              *         *         *         *         *         *
synlT.mze  AGCAACAGCAGCGTGAGCATCATCCGCGCCCCCATGTTCAGCTGGATCCACCGCAGCGCC
bssyn      ...........................T..A..T.............T........T...
synful.mod ...........................T..A..T.............T........T...

            1390      1400      1410      1420      1430      1440
              *         *         *         *         *         *
synlT.mze  GAGTTCAACAACATCATCCCCAGCAGCCAGATCACCCAGATCCCCCTGACCAAGAGCACC
bssyn      ............................................................
synful.mod ............................................................

            1450      1460      1470      1480      1490      1500
              *         *         *         *         *         *
synlT.mze  AACCTGGGCAGCGGCACCAGCGTGGTGAAGGGCCCCGGCTTCACCGGCGGCGACATCCTG
bssyn      ............................................................
synful.mod ............................................................

            1510      1520      1530      1540      1550      1560
              *         *         *         *         *         *
synlT.mze  CGCCGCACCAGCCCCGGCCAGATCAGCACCCTGCCCGTGAACATCACCGCCCCCCTGAGC
bssyn      ............................................................
synful.mod ............................................................

            1570      1580      1590      1600      1610      1620
              *         *         *         *         *         *
synlT.mze  CAGCGCTACCGCGTGCGCATCCGCTACGCCAGCACCACCAACCTGCAGTTCCACACCAGC
bssyn      ...............C............................................
synful.mod ...............C............................................

            1630      1640      1650      1660      1670      1680
              *         *         *         *         *         *
synlT.mze  ATCGACGGCCGCCCCCATCAACCAGGGCAACTTCAGCGCCACCATGAGCAGCGGCAGCAAC
bssyn      ............................................................
synful.mod ............................................................

            1690      1700      1710      1720      1730      1740
              *         *         *         *         *         *
synlT.mze  CTGCAGAGCGGCAGCTTCCGCACCGTGGGCTTCACCACCCCCTTCAACTTCAGCAACGGC
bssyn      ............................................................
synful.mod ............................................................

            1750      1760      1770      1780      1790      1800
              *         *         *         *         *         *
synlT.mze  AGCAGCGTGTTCACCCTGAGCGCCCACGTGTTCAACAGCGGCAACGAGGTGTACATCGAC
bssyn      ............................................................
synful.mod ............................................................
```

FIG. 3 (CONT)

```
           181J      1820      1830      184J      1850      1860
            *         *         *         *         *         *
synlT.mze  CGCATCGAGTTCGTGCCCGCCGAGGTGACCTTCGAGGCCGAGTACGACCTGGAGCGCGCC
bssyn      ..................................................A.G..T
synful.mod ..................................................A.G..T


           1870      1880      1890      1900      1910      1920
            *         *         *         *         *         *
synlT.mze  CAGAAGGCCGTGAACGAGCTGTTCACCAGCAGCAACCAGATCGGCCTGAAGACCGACGTG
bssyn      ............................................................
synful.mod ............................................................


           1930      1940      1950      1960      1970      1980
            *         *         *         *         *         *
synlT.mze  ACCGACTACCACATCGACCAGGTGAGCAACCTGGTGGAGTGCCTGAGCGACGAGTTCTGC
bssyn      ..................T.....------------------------------------
synful.mod ..................T.....------------------------------------


           1990      2000      2010      2020      2030      2040
            *         *         *         *         *         *
synlT.mze  CTGGACGAGAAGAAGGAGCTGAGCGAGAAGGTGAAGCACGCCAAGCGCCTGAGCGACGAG
bssyn      ------------------------------------------------------------
synful.mod ............................................................


           2050      2060      2070      2080      2090      2100
            *         *         *         *         *         *
synlT.mze  CGCAACCTGCTGCAGGACCCCAACTTCCGCGGCATCAACCGCCAGCTGGACCGCGGCTGG
bssyn      ------------------------------------------------------------
synful.mod ............................................................


           2110      2120      2130      2140      2150      2160
            *         *         *         *         *         *
synlT.mze  CGCGGCAGCACCGACATCACCATCCAGGGCGGCGACGACGTGTTCAAGGAGAACTACGTG
bssyn      ------------------------------------------------------------
synful.mod ............................................................


           2170      2180      2190      2200      2210      2220
            *         *         *         *         *         *
synlT.mze  ACCCTGCTGGGCACCTTCGACGAGTGCTACCCCACCTACCTGTACCAGAAGATCGACGAG
bssyn      ------------------------------------------------------------
synful.mod ............................................................


           2230      2240      2250      2260      2270      2280
            *         *         *         *         *         *
synlT.mze  AGCAAGCTGAAGGCCTACACCCGCTACCAGCTGCGCGGCTACATCGAGGACAGCCAGGAC
bssyn      ------------------------------------------------------------
synful.mod ............................................................


           2290      2300      2310      2320      2330      2340
            *         *         *         *         *         *
synlT.mze  CTGGAGATCTACCTGATCCGCTACAACGCCAAGCACGAGACCGTGAACGTGCCCGGCACC
bssyn      ------------------------------------------------------------
synful.mod ............................................................


           2350      2360      2370      2380      2390      2400
            *         *         *         *         *         *
synlT.mze  GGCAGCCTGTGGCCCCTGAGCGCCCCCAGCCCCATCGGCAAGTGCGCCCACCACAGCCAC
bssyn      ------------------------------------------------------------
synful.mod ............................................................
```

FIG. 3 (CONT)

```
             2410      2420      2430      2440      2450      2460
               *         *         *         *         *         *
synlT.mze  CACTTCAGCCTGGACATCGACGTGGGCTGCACCGACCTGAACGAGGACCTGGGCGTGTGG
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2470      2480      2490      2500      2510      2520
               *         *         *         *         *         *
synlT.mze  GTGATCTTCAAGATCAAGACCCAGGACGGCCACGCCCGCCTGGGCAACCTGGAGTTCCTG
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2530      2540      2550      2560      2570      2580
               *         *         *         *         *         *
synlT.mze  GAGGAGAAGCCCCTGGTGGGCGAGGCCCTGGCCCGCGTGAAGCGCGCCGAGAAGAAGTGG
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2590      2600      2610      2620      2630      2640
               *         *         *         *         *         *
synlT.mze  CGCGACAAGCGCGAGAAGCTGGAGTGGGAGACCAACATCGTGTACAAGGAGGCCAAGGAG
cssyn      ------------------------------------------------------------
synful.mod .........................................................

             2650      2660      2670      2680      2690      2700
               *         *         *         *         *         *
synlT.mze  AGCGTGGACGCCCTGTTCGTGAACAGCCAGTACGACCGCCTGCAGGCCGACACCAACATC
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2710      2720      2730      2740      2750      2760
               *         *         *         *         *         *
synlT.mze  GCCATGATCCACGCCGCCGACAAGCGCGTGCACAGCATCCGCGAGGCCTACCTGCCCGAG
bssyn      -------------------------------------------T----------------
synful.mod .........................................T...............

             2770      2780      2790      2800      2810      2820
               *         *         *         *         *         *
synlT.mze  CTGAGCGTGATCCCCGGCGTGAACGCCGCCATCTTCGAGGAGCTGGAGGGCCGCATCTTC
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2830      2840      2850      2860      2870      2880
               *         *         *         *         *         *
synlT.mze  ACCGCCTTCAGCCTGTACGACGCCCGCAACGTGATCAAGAACGGCGACTTCAACAACGGC
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2890      2900      2910      2920      2930      2940
               *         *         *         *         *         *
synlT.mze  CTGAGCTGCTGGAACGTGAAGGGCCACGTGGACGTGGAGGAGCAGAACAACCACCGCAGC
bssyn      ------------------------------------------------------------
synful.mod .........................................................

             2950      2960      2970      2980      2990      3000
               *         *         *         *         *         *
synlT.mze  GTGCTGGTGGTGCCCGAGTGGGAGGCCGAGGTGAGCCAGGAGGTGCGCGTGTGCCCCGGC
bssyn      ------------------------------------------------------------
synful.mod .........................................................
```

FIG. 3 (CONT)

```
          30..       3020       3030       304        3050       3060
            *          *          *          *          *          *
synlT.mze CGCGGCTACATCCTGCGCGTGACCGCCTACAAGGAGGGCTACGGCGAGGGCTGCGTGACC
bssyn     ------------------------------------------------------------
synful.mod .............................................................
```

```
          3070       3080       3090       3100       3110       3120
            *          *          *          *          *          *
synlT.mze ATCCACGAGATCGAGAACAACACCGACGAGCTGAAGTTCAGCAACTGCGTGGAGGAGGAG
bssyn     ------------------------------------------------------------
synful.mod .............................C...............................
```

```
          3130       3140       3150       3160       3170       3180
            *          *          *          *          *          *
synlT.mze GTGTACCCCAACAACACCGTGACCTGCAACGACTACACCGCCACCCAGGAGGAGTACGAG
bssyn     ------------------------------------------------------------
synful.mod .............................................................
```

```
          3190       3200       3210       3220       3230       3240
            *          *          *          *          *          *
synlT.mze GGCACCTACACCAGCCGCAACCGCGGCTACGACGGCGCCTACGAGAGCAACAGCAGCGTG
bssyn     ------------------------------------------------------------
synful.mod .............................................................
```

```
          3250       3260       3270       3280       3290       3300
            *          *          *          *          *          *
synlT.mze CCCGCCGACTACGCCAGCGCCTACGAGGAGAAGGCCTACACCGACGGCCGCCGCGACAAC
bssyn     ------------------------------------------------------------
synful.mod .............................................................
```

```
          3310       3320       3330       3340       3350       3360
            *          *          *          *          *          *
synlT.mze CCCTGCGAGAGCAACCGCGGCTACGGCGACTACACCCCCCCTGCCCGCCGGCTACGTGACC
bssyn     ------------------------------------------------------------
synful.mod .............................................................
```

```
          3370       3380       3390       3400       3410       3420
            *          *          *          *          *          *
synlT.mze AAGGAGCTGGAGTACTTCCCCGAGACCGACAAGGTGTGGATCGAGATCGGCGAGACCGAG
bssyn     ------------------------------------------------------------
synful.mod .............................................................
```

```
          3430       3440       3450       3460
            *          *          *          *
synlT.mze GGCACCTTCATCGTGGACAGCGTGGAGCTGCTGCTGATGGAGGAGTAG
bssyn     ------------------------------------------------....
synful.mod ..............................................
```

FIG. 4

```
          10        20        30       ¬0        50        60
           *         *         *         *         *         *
BTHKURHD  ATGGATAACAATCCGAACATCAATGAATGCATTCCTTATAATTGTTTAAGTAACCCTGAA
PMONBT    .....C.....C..A........C...........A..C..C..C..G........A...
bssyn     .....C.....C..C........C..G........C..C..C..C..CC.G..C....G

          70        80        90       100       110       120
           *         *         *         *         *         *
BTHKURHD  GTAGAAGTATTAGGTGGAGAAAGAATAGAAACTGGTTACACCCCAATCGATATTTCCTTG
PMONBT    ..T.......C.T..........C.C..T.....C.........T..C.....C..C....
bssyn     ..G..G..GC.G..C..C..GC.C..C..G..C..C........C.....C..CAG.C..

          130       140       150       160       170       180
           *         *         *         *         *         *
BTHKURHD  TCGCTAACGCAATTTCTTTTGAGTGAATTTGTTCCCGGTGCTGGATTTGTGTTAGGACTA
PMONBT    ..CT.G..A..G.....GC.C..C..G..C..G..A.......G..C..TC.C......
bssyn     AGC..G..C..G..C..GC....C..G..C..G.....C..C..C..C...C.G..C..G

          190       200       210       220       230       240
           *         *         *         *         *         *
BTHKURHD  GTTGATATAATATGGGGAATTTTTGGTCCCTCTCAATGGGACGCATTTCTTGTACAAATT
PMONBT    .....C..C..C.....T..C........A............T.....C..G..G......
bssyn     ..G..C..C..C.....C..C..C..C...AGC..G.......C..C..G..G..G..C

          250       260       270       280       290       300
           *         *         *         *         *         *
BTHKURHD  GAACAGTTAATTAACCAAAGAATAGAAGAATTCGCTAGGAACCAAGCCATTTCTAGATTA
PMONBT    ..G.....G..C.....G..G..C.....G.....C.........G.....C.....G..G
bssyn     ..G...C.G..C.....GC.C..C..G..G....CC.C.....G.....CAGCC.CC.G

          310       320       330       340       350       360
           *         *         *         *         *         *
BTHKURHD  GAAGGACTAAGCAATCTTTATCAAATTTACGCAGAATCTTTTAGAGAGTGGGAAGCAGAT
PMONBT    ......T.G........C..C.....C..T.....GAGC..C.............C....
bssyn     ..G..C..G.....C..G..C.....C.....C..GAGC..CC.C.........G..C..C

          370       380       390       400       410       420
           *         *         *         *         *         *
BTHKURHD  CCTACTAATCCAGCATTAAGAGAAGAGATGCGTATTCAATTCAATGACATGAACAGTGCC
PMONBT    ........C.....TC.CC.C..G..A.....................C..........C...
bssyn     ..C..C..C..C..CC.GC.C..G.........C..C..G....C...........C...

          430       440       450       460       470       480
           *         *         *         *         *         *
BTHKURHD  CTTACAACCGCTATTCCTCTTTTTGCAGTTCAAAATTATCAAGTTCCTCTTTTATCAGTA
PMONBT    T.G..C..A.....C..AT.G..C.....C..G..C..C...........C..G..C..G
bssyn     ..G..C.....C..C..C..G..C..C..G..G..C..C..G..G..C..GC.GAGC..G

          490       500       510       520       530       540
           *         *         *         *         *         *
BTHKURHD  TATGTTCAAGCTGCAAATTTACATTTATCAGTTTTGAGAGATGTTTCAGTGTTTGGACAA
PMONBT    ..C........A..T...C.T..CC.CAGC..GC.TC....C...AGC........G...
bssyn     ..C..G..G..C..C..CC.G..CC.GAGC..GC..C.C..C...CAGC.....C..G

          550       560       570       580       590       600
           *         *         *         *         *         *
BTHKURHD  AGGTGGGGATTTGATGCCGCGACTATCAATAGTCGTTATAATGATTTAACTAGGCTTATT
PMONBT    ...........C.....T..A..C........C.....C..C..CC.T........G...
bssyn     C.C.....C..C..C.....C..C........C..C..C..C..C..CC.G..CC.C..G..C
```

FIG. 4 (CONT)

```
            610       620       630       64U       650       660
             *         *         *         *         *         *
BTHKURHD   GGCAACTATACAGATCATGCTGTACGCTGGTACAATACGGGATTAGAGCGTGTATGGGGA
PMONBT     ..A.....C..C..C..C.....T..T........C..T..C..G.........C.....T
bssyn      ........C..C..C..C..C..G..........C..C..CC.G.....C..G....T

            670       680       690       700       710       720
             *         *         *         *         *         *
BTHKURHD   CCGGATTCTAGAGATTGGATAAGATATAATCAATTTAGAAGAGAATTAACACTAACTGTA
PMONBT     ..T................T.....C..C..G..C..G.........G..C..C..A..T
bssyn      ..C..CAGCC.C..C.....C..G..C..C..G..CC.CC.C..GC.G..C..G..C..G

            730       740       750       760       770       780
             *         *         *         *         *         *
BTHKURHD   TTAGATATCGTTTCTCTATTTCCGAACTATGATAGTAGAACGTATCCAATTCGAACAGTT
PMONBT     ..G..C..T..G.....C..C..........CTCC.....C..C..T..C..T.....G
bssyn      C.G..C.....GAGC..G..C..C.....C..C..CC.C..C..C..C..C..C..C..G

            790       800       810       820       830       840
             *         *         *         *         *         *
BTHKURHD   TCCCAATTAACAAGAGAAATTTATACAAACCCAGTATTAGAAAATTTTGATGGTAGTTTT
PMONBT     ......C.T..C.........C.....T......TC.T..G..C..C..C.....C..C
bssyn      AG...GC.G..CC.C..G.....C..C.....C..GC.G..G..C..C..C..C..C

            350       360       870       880       590       900
             *         *         *         *         *         *
BTHKURHD   CGAGGCTCGGCTCAGGGCATAGAAGGAAGTATTAGGAGTCCACATTTGATGGATATACTT
PMONBT     ..T..T..T..C..A..T..C.....CTCC..C.....C.....C..........C..CT.G
bssyn      ..C...AGC..C........C..G..C..C..CC.C..C..C..CC.......C..C..G

            910       920       930       940       950       960
             *         *         *         *         *         *
BTHKURHD   AACAGTATAACCATCTATACGGATGCTCATAGAGGAGAATATTATTGGTCAGGGCATCAA
PMONBT     .....C.....T.....C.GC........C.......G......C.....C...T..A..C..G
bssyn      .....C..C.....C..C..C..C..CC.C..C..G..C...AGC..C..C..G

            970       980       990       1000      1010      1020
             *         *         *         *         *         *
BTHKURHD   ATAATGGCTTCTCCTGTAGGGTTTTCGGGGCCAGAATTCACTTTTCCGCTATATGGAACT
PMONBT     ..C.....C.....A..T..A..CAGC.....C..G..T..C.....T..C.........
bssyn      ..C.....CAGC..C..C..C..CAGC..C..C..G.....C..C..C..G..C..C..C

            1030      1040      1050      1060      1070      1080
             *         *         *         *         *         *
BTHKURHD   ATGGGAAATGCAGCTCCACAACAACGTATTGTTGCTCAACTAGGTCAGGGCGTGTATAGA
PMONBT     ........C..C..............................C...............T..C..C...
bssyn      .....C..C..T..A..T..G..G..C..C..G..A..G..G..C.....A.....CC.C

            1090      1100      1110      1120      1130      1140
             *         *         *         *         *         *
BTHKURHD   ACATTATCGTCCACTTTATATAGAAGACCTTTTAATATAGGGATAAATAATCAACAACTA
PMONBT     ..C..G..T.....C..G..C.........C..C.....C..T..C..C..C..G....T
bssyn      ..CC.GAGCAG...CC.G..CC.TC.......C..C..C..C..C..C..C..G..G..G

            1150      1160      1170      1180      1190      1200
             *         *         *         *         *         *
BTHKURHD   TCTGTTCTTGACGGGACAGAATTTGCTTATGGAACCTCCTCAAATTTGCCATCCGCTGTA
PMONBT     ..C...........A.....G..C..C...........T..T..C..............T
bssyn      AGC..G..G.....C..C..G..C..C..C..C...AG.AGC..CC....CAG...C..G
```

FIG. 4 (CONT)

```
           12.0      1220      1230      124.      1250      1260
            *         *         *         *         *         *
BTHKURHD  TACAGAAAAAGCGGAACGGTAGATTCGCTGGATGAAATACCGCCACAGAATAACAACGTG
PMONBT    .......G........C..T.....CT....C.....C..A........C.....T...
bssyn     ...C.C..G.....C..C..G..CAGC.....C..G..C..C..T.....C........


           1270      1280      1290      1300      1310      1320
            *         *         *         *         *         *
BTHKURHD  CCACCTAGGCAAGGATTTAGTCATCGATTAAGCCATGTTTCAATGTTTCGTTCAGGCTTT
PMONBT    .....C...........CTCC..CA.G..G.....C..G..C.....C.....C..A..C
bssyn     ......C.A..G..C..C..C..C..TC.G.....C..GAGC.....C..CAGT.....C


           1330      1340      1350      1360      1370      1380
            *         *         *         *         *         *
BTHKURHD  AGTAATAGTAGTGTAAGTATAATAAGAGCTCCTATGTTCTCTTGGATACATCGTAGTGCT
PMONBT    ..C..C...TCC..G..C..C..C.................A.....T...........
bssyn     ..C..C..C..G..C..C..CC.T..A.........AGC.....T..C..C.....C


           1390      1400      1410      1420      1430
            *         *         *         *         *
BTHKURHD  GAATTTAATAATATAATTCCTTCATCA--CAAATTACACAAATACCTTTAACAAAATCTA
PMONBT    ..G..C..C.....C........C..T--.....C..C.....C..A..G..C..G....
bssyn     ..G..C..C..C..C..C..---..G..GC..G..C..C..G..C..CC.G..C..GAGC.


           1440      1450      1460      1470      1480      1490
            *         *         *         *         *         *
BTHKURHD  CTAATCTTGGCTCTGGAACTTCTGTCGTTAAAGGACCAGGATTTACAGGAGGAGATATTC
PMONBT    ....C.....A................G.........C..C........T.......
bssyn     .C..C..G...AGC..C..CAGC..G..G..G..C..C..C..C..C..C..C..C..C.


          1500      1510      1520      1530      1540      1550
            *         *         *         *         *         *
BTHKURHD  TTCGAAGAACTTCACCTGGCCAGATTTCAACCTTAAGAGTAAATATTACTGCACCATTAT
PMONBT    ..A...........T.........AGC...C.C.....T..C..C.........C.T.
bssyn     .G..CC.C..CAGC..C.........CAGC...C.GC.C..G..C..C..C..C..CC.GA


          1560      1570      1580      1590      1600      1610
            *         *         *         *         *         *
BTHKURHD  CACAAAGATATCGGGTAAGAATTCGCTACGCTTCTACCACAAATTTACAATTCCATACAT
PMONBT    .T..........T..C..G.....T.....A.......T..C..G.........C..C.
bssyn     GC..GC.C..C..C..CC.C..C........CAGC.....C..CC.G..G.....C..CA


          1620      1630      1640      1650      1660      1670
            *         *         *         *         *         *
BTHKURHD  CAATTGACGGAAGACCTATTAATCAGGGGAATTTTTTCAGCAACTATGAGTAGTGGGAGTA
PMONBT    .C..C........G.....C..........T..C..C..C.....C...TCA..C..C..C.
bssyn     GC..C.....CC.C..C..C..C.....C..C..CAGC..C..C.....C..C..C..C.


          1680      1690      1700      1710      1720      1730
            *         *         *         *         *         *
BTHKURHD  ATTTACAGTCCGGAAGCTTTAGGACTGTAGGTTTTACTACTCCGTTTAACTTTTCAAATG
PMONBT    .C..G..A.....C.....C..A..C..C.....C........T..C.....C..T..C.
bssyn     .CC.G...AG...C.....CC.C..C..G..C..C..C..C..C.....CAGC..C.


          1740      1750      1760      1770      1780      1790
            *         *         *         *         *         *
BTHKURHD  GATCAAGTGTATTTACGTTAAGTGCTCATGTCTTCAATTCAGGCAATGAAGTTTATATAG
PMONBT    ......C..T..C..CC.T..C.........G........T...........G..C..T.
bssyn     .CAGC..C..G..C..CC.G..C..C..C..G.....CAGC.....C..G..G..C..C.
```

18/94

FIG. 4 (CONT)

```
           1800        310        1820       1830        .40
            *           *          *          *           *
BTHKURHD   ATCGAATTGAATTTGTTCCGGCAGAAGTAACCTTTGAGGCAGAATA--------------
PMONBT     .C..T.....G.....G..T..C.....T.....C.....T..G..--------------
bssyn      .C..C..C..G..C..G..C..C..G..G.....C.....C..G..CGACCTGGAGAGGG


BTHKURHD   -------------------------------------------------------------
PMONBT     -------------------------------------------------------------
bssyn      CTCAGAAGGCCGTGAACGAGCTGTTCACCAGCAGCAACCAGATCGGCCTGAAGACCGACG


BTHKURHD   -------------------------------T
PMONBT     -------------------------------C
bssyn      TGACCGACTACCACATCGATCAGGTGTAG
```

FIG. 5

```
               10        20        30        .0        50        60
                *         *         *         *         *         *
PHONBT    ATGGACAACAACCCAAACATCAACGAATGCATTCCATACAACTGCTTGAGTAACCCAGAA
bssyn     ..............C..........G........C..C.........C....C..G

               70        80        90        100       110       120
                *         *         *         *         *         *
PMONBT    GTTGAAGTACTTGGTGGAGAACGCATTGAAACCGGTTACACTCCCATCGACATCTCCTTG
bssyn     ..G..G..G..G..C..C..G.....C..G.....C.....C...........AG.C..

               130       140       150       160       170       180
                *         *         *         *         *         *
PMONBT    TCCTTGACACAGTTTCTGCTCAGCGAGTTCGTGCCAGGTGCTGGGTTCGTTCTCGGACTA
bssyn     AG.C.....C.....C.....G.........C..C..C..C.....G..G..C..G

               190       200       210       220       230       240
                *         *         *         *         *         *
PMONBT    GTTGACATCATCTGGGGTATCTTTGGTCCATCTCAATGGGATGCATTCCTGGTGCAAATT
bssyn     ..G.............C........C..C..CAGC..G.....C..C..........G..C

               250       260       270       280       290       300
                *         *         *         *         *         *
PMONBT    GAGCAGTTGATCAACCAGAGGATCGAAGAGTTCGCCAGGAACCAGGCCATCTCTAGGTTG
bssyn     ......C...........C.C.....G.........C.C...............AGCC.CC..

               310       320       330       340       350       360
                *         *         *         *         *         *
PMONBT    GAAGGATTGAGCAATCTCTACCAAATCTATGCAGAGAGCTTCAGAGAGTGGGAAGCCGAT
bssyn     ..G..CC.......C..G...........C..C.........C.C........G.....C

               370       380       390       400       410       420
                *         *         *         *         *         *
PMONBT    CCTACTAACCCAGCTCTCCGCGAGGAAATGCGTATTCAATTCAACGACATGAACAGCGCC
bssyn     ..C..C.....C..C..G........G.....C..C..G...................

               430       440       450       460       470       480
                *         *         *         *         *         *
PMONBT    TTGACCACAGCTATCCCATTGTTCGCAGTCCAGAACTACCAAGTTCCTCTCTTGTCCGTG
bssyn     C.......C..C.....CC.......C..G...........G..G..C..GC..AG....

               490       500       510       520       530       540
                *         *         *         *         *         *
PMONBT    TACGTTCAAGCAGCTAATCTTCACCTCAGCGTGCTTCGAGACGTTAGCGTGTTTGGGCAA
bssyn     .....G..G..C..C..C..G.....G........G..C.....C.........C..C..G

               550       560       570       580       590       600
                *         *         *         *         *         *
PMONBT    AGGTGGGGATTCGATGCTGCAACCATCAATAGCCGTTACAACGACCTTACTAGGCTGATT
bssyn     C.C.....C.....C..C..C.........C.....C.............G..CC.C.....C

               610       620       630       640       650       660
                *         *         *         *         *         *
PMONBT    GGAAACTACACCGACCACGCTGTTCGTTGGTACAACACTGGCTTGGAGCGTGTCTGGGGT
bssyn     ..C.................C..G..C..........C...........C..G......

               670       680       690       700       710       720
                *         *         *         *         *         *
PMONBT    CCTGATTCTAGAGATTGGATTAGATACAACCAGTTCAGGAGAGAATTGACCCTCACAGTT
bssyn     ..C..CAGCC.C..C.....C..G............C.CC.C..GC.......G..C..G
```

20/94

FIG. 5 (CONT)

```
         0      740       750      7bJ       770       780
         *       *         *        *         *         *
PMONBT  TTGGACATTGTGTCTCTCTTCCCGAACTATGACTCCAGAACCTACCCTATCCGTACAGTG
bssyn   C.......C...AGC..G.....C.....C...AG.C.C........C.....C..C...


        790       800       810       820       830       840
         *         *         *         *         *         *
PMONBT  TCCCAACTTACCAGAGAAATCTATACTAACCCAGTTCTTGAGAACTTCGACGGTAGCTTC
bssyn   AG...G..G...C.C..G..T..C..C.....C..G..G..............C......


        850       860       870       880       890       900
         *         *         *         *         *         *
PMONBT  CGTGGTTCTGCCCAAGGTATCGAAGGCTCCATCAGGAGCCCACACTTGATGGACATCTTG
bssyn   ..C..CAGC.....G..C.....G...AG....C.C.....C...C..........C..


        910       920       930       940       950       960
         *         *         *         *         *         *
PMONBT  AACAGCATAACTATCTACAGCGATGCTCACAGAGGAGAGTATTACTGGTCTGGACACCAG
bssyn   ........C..C.......C..C..C...C.C..C.....C......AGC..C......


        970       980       990      1000      1010      1020
         *         *         *         *         *         *
PMONBT  ATCATGGCCTCTCCAGTTGGATTCAGCGGGCCCGAGTTTACCTTTCCTCTCTATGGAACT
bssyn   .........AGC..C..C..C........C........C.....C..C..G..C..C..C


       1030      1040      1050      1060      1070      1080
         *         *         *         *         *         *
PMONBT  ATGGGAAACGCCGCTCCACAACAACGTATCGTTGCTCAACTAGGTCAGGGTGTCTACAGA
bssyn   .....C.....T..A..T..G..G..C.....G..A..G..G..C.....A..G...C.C


       1090      1100      1110      1120      1130      1140
         *         *         *         *         *         *
PMINET  ACCTTGTCTTCCACCTTGTACAGAAGACCCTTCAATATCGGTATCAACAACCAGCAACTT
bssyn   ...C..AGCAG....C.....C.TC....T.....C.....C..............G..G


       1150      1160      1170      1180      1190      1200
         *         *         *         *         *         *
PMONBT  TCCGTTCTTGACGGAACAGAGTTCGCCTATGGAACCTCTTCTAACTTGCCATCCGCTGTT
bssyn   AG...G..G.....C..C...........C..C...AGCAGC...C....CAG...C..G


       1210      1220      1230      1240      1250      1260
         *         *         *         *         *         *
PMONBT  TACAGAAAGAGCGGAACCGTTGATTCCTTGGACGAAATCCCACCACAGAACAACAATGTG
bssyn   ...C.C........C.....G..CAG.C.......G.....C..T..........C...


       1270      1280      1290      1300      1310      1320
         *         *         *         *         *         *
PMONBT  CCACCCAGGCAAGGATTCTCCCACAGGTTGAGCCACGTGTCCATGTTCCGTTCCGGATTC
bssyn   .....TC.A..G..C...AG....C.TC............AG.........CAGT..C...


       1330      1340      1350      1360      1370      1380
         *         *         *         *         *         *
PMONBT  AGCAACAGTTCCGTGAGCATCATCAGAGCTCCTATGTTCTCATGGATTCATCGTAGTGCT
bssyn   ........CAG.............C.T..A.........AGC.........C..C.....C


       1390      1400      1410      1420      1430      1440
         *         *         *         *         *         *
PMONBT  GAGTTCAACAATATCATTCCTTCCTCTCAAATCACCCAAATCCCATTGACCAAGTCTACT
bssyn   ..........C.....C..CAG.AGC..G........G.....CC..........AGC..C
```

FIG. 5 (CONT)

```
         14.0      1460      1470      146v      1490      1500
           *         *         *         *         *         *
PHONBT   AACCTTGGATCTGGAACTTCTGTCGTGAAAGGACCAGGCTTCACAGGAGGTGATATTCTT
bssyn    .....G..CAGC...C..CAGC..G.....G..C..C........C..C..C..C..C..G

         1510      1520      1530      1540      1550      1560
           *         *         *         *         *         *
PHONBT   AGAAGAACTTCTCCTGGCCAGATTAGCACCCTCAGAGTTAACATCACTGCACCACTTTCT
pssyn    C.CC.C..CAGC..C........C........GC.C..G........C..C..C..GAGC

           ..  1570      1580      1590      1600      1610      1620
                  *         *         *         *         *         *
PHONBT   CAAAGATATCGTGTCAGGATTCGTTACGCATCTACCACTAACTTGCAATTCCACACCTCC
bssyn    ..GC.C..C..C...C.C..C..C.....CAGC.....C...C....G.........AG.

         1630      1640      1650      1660      i670      1680
           *         *         *         *         *         *
PHONBT   ATCGACGGAAGGCCTATCAATCAGGGTAACTTCTCCGCAACCATGTCAAGCGGCAGCAAC
bssyn    ........CC.C..C.....C.....C.....AG...C......AGC..........

         1690      1700      1710      1720      1730      1740
           *         *         *         *         *         *
FMONBT   TTGCAATCCGGCAGCTTCAGAACCGTCGGTTTCACTACTCCTTTCAACTTCTCTAACGGA
bssyn    C....GAG.........C.C.....G..C.....C..C..C.........AGC.....C

         1750      1760      1770      1780      1790      1800
           *         *         *         *         *         *
FMONBT   TCAAGCGTTTTCACCCTTAGCGCTCATGTGTTCAATTCTGGCAATGAAGTGTACATTGAC
bssyn    AGC.....G........G.....C..C.........CAGC.....C..G........C...

         1810      1820      1830      1840
           *         *         *         *
FMONBT   CGTATTGAGTTTGTGCCTGCCGAAGTTACCTTCGAGGCTGAGTA----------------
bssyn    ..C..C.....C.....C.....G..G............C.....CGACCTGGAGAGGGCT


FMONBT   -------------------------------------------------------------
bssyn    CAGAAGGCCGTGAACGAGCTGTTCACCAGCAGCAACCAGATCGGCCTGAAGACCGACGTG


FMONBT   --------------------------C
bssyn    ACCGACTACCACATCGATCAGGTGTAG
```

FIG. 6

```
  64  ATGGACCTGC TGCCCGACGC CCGCATCGAG GACAGCCTGT GCATCGCCGA GGGCAACAAC
      MetAspLeu LeuProAsp AlaArgIleGlu AspSerLeu CysIleAla GluGlyAsnAsn

 124  ATCGACCCCT TCGTGAGCGC CAGCACCGTG CAGACCGGCA TCAACATCGC CGGCCGCATC
      IleAspPro PheValSer AlaSerThrVal GlnThrGly IleAsnIle AlaGlyArgIle

 184  CTGGGCGTGC TGGGCGTGCC CTTCGCCGGC CAGCTGGCCA GCTTCTACAG CTTCCTGGTG
      LeuGlyVal LeuGlyVal ProPheAlaGly GlnLeuAla SerPheTyr SerPheLeuVal

 244  GGCGAGCTGT GGCCCCGCGG CCGCGACCAG TGGGAGATCT TCCTGGAGCA CGTGGAGCAG
      GlyGluLeu TrpProArg GlyArgAspGln TrpGluIle PheLeuGlu HisValGluGln

 304  CTGATCAACC AGCAGATCAC CGAGAACGCC CGCAACACCG CCCTGGCCCG CCTGCAGGGC
      LeuIleAsn GlnGlnIle ThrGluAsnAla ArgAsnThr AlaLeuAla ArgLeuGlnGly

 364  CTGGGCGACA GCTTCCGCGC CTACCAGCAG AGCCTGGAGG ACTGGCTGGA GAACCGCGAC
      LeuGlyAsp SerPheArg AlaTyrGlnGln SerLeuGlu AspTrpLeu GluAsnArgAsp

 424  GACGCCCGCA CCCGCAGCGT GCTGTACACC CAGTACATCG CCCTGGAGCT GGACTTCCTG
      AspAlaArg ThrArgSer ValLeuTyrThr GlnTyrIle AlaLeuGlu LeuAspPheLeu

 484  AACGCCATGC CCCTGTTCGC CATCCGCAAC CAGGAGGTGC CCCTGCTGAT GGTGTACGCC
      AsnAlaMet ProLeuPhe AlaIleArgAsn GlnGluVal ProLeuLeu MetValTyrAla

 544  CAGGCCGCCA ACCTGCACCT GCTGCTGCTG CGCGACGCCA GCCTGTTCGG CAGCGAGTTC
      GlnAlaAla AsnLeuHis LeuLeuLeuLeu ArgAspAla SerLeuPhe GlySerGluPhe

 604  GGCCTGACCA GCCAGGAGAT CCAGCGCTAC TACGAGCGCC AGGTGGAGCG CACCCGCGAC
      GlyLeuThr SerGlnGlu IleGlnArgTyr TyrGluArg GlnValGlu ArgThrArgAsp

 664  TACAGCGACT ACTGCGTGGA GTGGTACAAC ACCGGCCTGA ACAGCCTGCG CGGCACCAAC
      TyrSerAsp TyrCysVal GluTrpTyrAsn ThrGlyLeu AsnSerLeu ArgGlyThrAsn

 724  GCCGCCAGCT GGGTGCGCTA CAACCAGTTC CGCCGCGACC TGACCCTGGG CGTGCTGGAC
      AlaAlaSer TrpValArg TyrAsnGlnPhe ArgArgAsp LeuThrLeu GlyValLeuAsp

 784  CTGGTGGCCC TGTTCCCCAG CTACGACACC CGCACCTACC CCATCAACAC CAGCGCCCAG
      LeuValAla LeuPhePro SerTyrAspThr ArgThrTyr ProIleAsn ThrSerAlaGln

 844  CTGACCCGCG AGGTGTACAC CGACGCCATC GGCGCCACCG GCGTGAACAT GGCCAGCATG
      LeuThrArg GluValTyr ThrAspAlaIle GlyAlaThr GlyValAsn MetAlaSerMet

 904  AACTGGTACA ACAACAACGC CCCCAGCTTC AGCGCCATCG AGGCCGCCGC CATCCGCAGC
      AsnTrpTyr AsnAsnAsn AlaProSerPhe SerAlaIle GluAlaAla AlaIleArgSer

 964  CCCCACCTGC TGGACTTCCT GGAGCAGCTG ACCATCTTCA GCGCCAGCAG CCGCTGGAGC
      ProHisLeu LeuAspPhe LeuGluGlnLeu ThrIlePhe SerAlaSer SerArgTrpSer

1024  AACACCCGCC ACATGACCTA CTGGCGCGGC CACACCATCC AGAGCCGCCC CATCGGCGGC
      AsnThrArg HisMetThr TyrTrpArgGly HisThrIle GlnSerArg ProIleGlyGly
```

FIG. 6 (CONT)

```
1084   GGCCTGAACA CCAGCACCCA CGGCGCCACC AACACCAGCA TCAACCCCGT GACCCTGCGC
       GlyLeuAsn ThrSerThr HisGlyAlaThr AsnThrSer IleAsnPro ValThrLeuArg

1144   TTCGCCAGCC GCGACGTGTA CCGCACCGAG AGCTACGCCG GCGTGCTGCT GTGGGGCATC
       PheAlaSer ArgAspVal TyrArgThrGlu SerTyrAla GlyValLeu LeuTrpGlyIle

1204   TACCTGGAGC CCATCCACGG CGTGCCCACC GTGCGCTTCA ACTTCACCAA CCCCCAGAAC
       TyrLeuGlu ProIleHis GlyValProThr ValArgPhe AsnPheThr AsnProGlnAsn

1264   ATCAGCGACC GCGGCACCGC CAACTACAGC CAGCCCTACG AGAGCCCCGG CCTGCAGCTG
       IleSerAsp ArgGlyThr AlaAsnTyrSer GlnProTyr GluSerPro GlyLeuGlnLeu

1324   AAGGACAGCG AGACCGAGCT GCCCCCCGAG ACCACCGAGC GCCCCAACTA CGAGAGCTAC
       LysAspSer GluThrGlu LeuProProGlu ThrThrGlu ArgProAsn TyrGluSerTyr

1384   AGCCACCGCC TGAGCCACAT CGGCATCATC CTGCAGAGCC GCGTGAACGT GCCCGTGTAC
       SerHisArg LeuSerHis IleGlyIleIle LeuGlnSer ArgValAsn ValProValTyr

1444   AGCTGGACCC ACCGCAGCGC CGACCGCACC AACACCATCG GCCCCAACCG CATCACCCAG
       SerTrpThr HisArgSer AlaAspArgThr AsnThrIle GlyProAsn ArgIleThrGln

1504   ATCCCCATGG TGAAGGCCAG CGAGCTGCCC CAGGGCACCA CCGTGGTGCG CGGCCCCGGC
       IleProMet ValLysAla SerGluLeuPro GlnGlyThr ThrValVal ArgGlyProGly

1564   TTCACCGGCG GCGACATCCT GCGCCGCACC AACACCGGCG GCTTCGGCCC CATCCGCGTG
       PheThrGly GlyAspIle LeuArgArgThr AsnThrGly GlyPheGly ProIleArgVal

1624   ACCGTGAACG GCCCCCTGAC CCAGCGCTAC CGCATCGGCT TCCGCTACGC CAGCACCGTG
       ThrValAsn GlyProLeu ThrGlnArgTyr ArgIleGly PheArgTyr AlaSerThrVal

1684   GACTTCGACT TCTTCGTGAG CCGCGGCGGC ACCACCGTGA CAACTTCCG CTTCCTGCGC
       AspPheAsp PhePheVal SerArgGlyGly ThrThrVal AsnAsnPhe ArgPheLeuArg

1744   ACCATGAACA GCGGCGACGA GCTGAAGTAC GGCAACTTCG TGCGCCGCGC CTTCACCACC
       ThrMetAsn SerGlyAsp GluLeuLysTyr GlyAsnPhe ValArgArg AlaPheThrThr

1804   CCCTTCACCT TCACCCAGAT CCAGGACATC ATCCGCACCA GCATCCAGGG CCTGAGCGGC
       ProPheThr PheThrGln IleGlnAspIle IleArgThr SerIleGln GlyLeuSerGly

1864   AACGGCGAGG TGTACATCGA CAAGATCGAG ATCATCCCCG TGACCGCCAC CTTCGAGGCC
       AsnGlyGlu ValTyrIle AspLysIleGlu IleIlePro ValThrAla ThrPheGluAla

1924   GAGTACGACC TGGAGCGCGC CCAGGAGGCC GTGAACGCCC TGTTCACCAA CACCAACCCC
       GluTyrAsp LeuGluArg AlaGlnGluAla ValAsnAla LeuPheThr AsnThrAsnPro

1984   CGCCGCCTGA AGACCGACGT GACCGACTAC CACATCGACC AGGTGAGCAA CCTGGTGGCC
       ArgArgLeu LysThrAsp ValThrAspTyr HisIleAsp GlnValSer AsnLeuValAla

2044   TGCCTGAGCG ACGAGTTCTG CCTGGACGAG AAGCGCGAGC TGCTGGAGAA GGTGAAGTAC
       CysLeuSer AspGluPhe CysLeuAspGlu LysArgGlu LeuLeuGlu LysValLysTyr
```

FIG. 6 (CONT)

2104   GCCAAGCGCC TGAGCGACGA GCGCAACCTG CTGCAGGACC CCAACTTCAC CAGCATCAAC
          AlaLysArg LeuSerAsp GluArgAsnLeu LeuGlnAsp ProAsnPhe ThrSerIleAsn

2164   AAGCAGCCCG ACTTCATCAG CACCAACGAG CAGAGCAACT TCACCAGCAT CCACGAGCAG
          LysGlnPro AspPheIle SerThrAsnGlu GlnSerAsn PheThrSer IleHisGluGln

2224   AGCGAGCACG GCTGGTGGGG CAGCGAGAAC ATCACCATCC AGGAGGGCAA CGACGTGTTC
          SerGluHis GlyTrpTrp GlySerGluAsn IleThrIle GlnGluGly AsnAspValPhe

2284   AAGGAGAACT ACGTGACCCT GCCCGGCACC TTCAACGAGT GCTACCCCAC CTACCTGTAC
          LysGluAsn TyrValThr LeuProGlyThr PheAsnGlu CysTyrPro ThrTyrLeuTyr

2344   CAGAAGATCG GCGAGAGCGA GCTGAAGGCC TACACCCGCT ACCAGCTGCG CGGCTACATC
         GlnLysIle GlyGluSer GluLeuLysAla TyrThrArg TyrGlnLeu ArgGlyTyrIle

2404   GAGGACAGCC AGGACCTGGA GATCTACCTG ATCCGCTACA ACGCCAAGCA CGAGACCCTG
          GluAspSer GlnAspLeu GluIleTyrLeu IleArgTyr AsnAlaLys HisGluThrLeu

2464   GACGTGCCCG GCACCGAGAG CCTGTGGCCC CTGAGCGTGG AGAGCCCCAT CGGCCGCTGC
          AspValPro GlyThrGlu SerLeuTrpPro LeuSerVal GluSerPro IleGlyArgCys

2524   GGCGAGCCCA ACCGCTGCGC CCCCCACTTC GAGTGGAACC CCGACCTGGA CTGCAGCTGC
          GlyGluPro AsnArgCys AlaProHisPhe GluTrpAsn ProAspLeu AspCysSerCys

2584   CGCGACGGCG AGAAGTGCGC CCACCACAGC CACCACTTCA GCCTGGACAT CGACGTGGGC
          ArgAspGly GluLysCys AlaHisHisSer HisHisPhe SerLeuAsp IleAspValGly

2644   TGCACCGACC TGCACGAGAA CCTGGGCGTG TGGGTGGTGT TCAAGATCAA GACCCAGGAG
          CysThrAsp LeuHisGlu AsnLeuGlyVal TrpValVal PheLysIle LysThrGlnGlu

2704   GGCCACGCCC GCCTGGGCAA CCTGGAGTTC ATCGAGGAGA GCCCCTGCT GGGCGAGGCC
          GlyHisAla ArgLeuGly AsnLeuGluPhe IleGluGlu LysProLeu LeuGlyGluAla

2764   CTGAGCCGCG TGAAGCGCGC CGAGAAGAAG TGGCGCGACA AGCGCGAGAA GCTGCAGCTG
          LeuSerArg ValLysArg AlaGluLysLys TrpArgAsp LysArgGlu LysLeuGlnLeu

2824   GAGACCAAGC GCGTGTACAC CGAGGCCAAG GAGGCCGTGG ACGCCCTGTT CGTGGACAGC
          GluThrLys ArgValTyr ThrGluAlaLys GluAlaVal AspAlaLeu PheValAspSer

2884   CAGTACGACC GCCTGCAGGC CGACACCAAC ATCGGCATGA TCCACGCCGC CGACAAGCTG
          GlnTyrAsp ArgLeuGln AlaAspThrAsn IleGlyMet IleHisAla AlaAspLysLeu

2944   GTGCACCGCA TCCGCGAGGC CTACCTGAGC GAGCTGCCCG TGATCCCCGG CGTGAACGCC
          ValHisArg IleArgGlu AlaTyrLeuSer GluLeuPro ValIlePro GlyValAsnAla

3004   GAGATCTTCG AGGAGCTGGA GGGCCACATC ATCACCGCCA TCAGCCTGTA CGACGCCCGC
          GluIlePhe GluGluLeu GluGlyHisIle IleThrAla IleSerLeu TyrAspAlaArg

FIG 6 (CONT)

3064 AACGTGGTGA AGAACGGCGA CTTCAACAAC GGCCTGACCT GCTGGAACGT GAAGGGCCAC
      AsnValVal LysAsnGly AspPheAsnAsn GlyLeuThr CysTrpAsn ValLysGlyHis

3124 GTGGACGTGC AGCAGAGCCA CCACCGCAGC GACCTGGTGA TCCCCGAGTG GGAGGCCGAG
      ValAspVal GlnGlnSer HisHisArgSer AspLeuVal IleProGlu TrpGluAlaGlu

3184 GTGAGCCAGG CCGTGCGCGT GTGCCCCGGC TGCGGCTACA TCCTGCGCGT GACCGCCTAC
      ValSerGln AlaValArg ValCysProGly CysGlyTyr IleLeuArg ValThrAlaTyr

3244 AAGGAGGGCT ACGGCGAGGG CTGCGTGACC ATCCACGAGA TCGAGAACAA CACCGACGAG
      LysGluGly TyrGlyGlu GlyCysValThr IleHisGlu IleGluAsn AsnThrAspGlu

3304 CTGAAGTTCA AGAACCGCGA GGAGGAGGAG GTGTACCCCA CCGACACCGG CACCTGCAAC
      LeuLysPhe LysAsnArg GluGluGluGlu ValTyrPro ThrAspThr GlyThrCysAsn

3364 GACTACACCG CCCACCAGGG CACCGCCGGC TGCGCCGACG CCTGCAACAG CCGCAACGCC
      AspTyrThr AlaHisGln GlyThrAlaGly CysAlaAsp AlaCysAsn SerArgAsnAla

3424 GGCTACGAGG ACGCCTACGA GGTGGACACC ACCGCCAGCG TGAACTACAA GCCCACCTAC
      GlyTyrGlu AspAlaTyr GluValAspThr ThrAlaSer ValAsnTyr LysProThrTyr

3484 GAGGAGGAGA CCTACACCGA CGTGCGCCGC GACAACCACT GCGAGTACGA CCGCGGCTAC
      GluGluGlu ThrTyrThr AspValArgArg AspAsnHis CysGluTyr AspArgGlyTyr

3544 GTGAACTACC CCCCCGTGCC CGCCGGCTAC GTGACCAAGG AGCTGGAGTA CTTCCCCGAG
      ValAsnTyr ProProVal ProAlaGlyTyr ValThrLys GluLeuGlu TyrPheProGlu

3604 ACCGACACCG TGTGGATCGA GATCGGCGAG ACCGAGGGCA AGTTCATCGT GGACAGCGTG
      ThrAspThr ValTrpIle GluIleGlyGlu ThrGluGly LysPheIle ValAspSerVal

3664 GAGCTGCTGC TGATGGAGGA GTAG
      GluLeuLeu LeuMetGlu Glu---

26/94

FIG. 6

```
  64   ATGGACCTGC TGCCCGACGC CCGCATCGAG GACAGCCTGT GCATCGCCGA GGGCAACAAC
       MetAspLeu LeuProAsp AlaArgIleGlu AspSerLeu CysIleAla GluGlyAsnAsn

 124   ATCGACCCCT TCGTGAGCGC CAGCACCGTG CAGACCGGCA TCAACATCGC CGGCCGCATC
       IleAspPro PheValSer AlaSerThrVal GlnThrGly IleAsnIle AlaGlyArgIle

 184   CTGGGCGTGC TGGGCGTGCC CTTCGCCGGC CAGCTGGCCA GCTTCTACAG CTTCCTGGTG
       LeuGlyVal LeuGlyVal ProPheAlaGly GlnLeuAla SerPheTyr SerPheLeuVal

 244   GGCGAGCTGT GGCCCCGCGG CCGCGACCAG TGGGAGATCT TCCTGGAGCA CGTGGAGCAG
       GlyGluLeu TrpProArg GlyArgAspGln TrpGluIle PheLeuGlu HisValGluGln

 304   CTGATCAACC AGCAGATCAC CGAGAACGCC CGCAACACCG CCCTGGCCCG CCTGCAGGGC
       LeuIleAsn GlnGlnIle ThrGluAsnAla ArgAsnThr AlaLeuAla ArgLeuGlnGly

 364   CTGGGCGACA GCTTCCGCGC CTACCAGCAG AGCCTGGAGG ACTGGCTGGA GAACCGCGAC
       LeuGlyAsp SerPheArg AlaTyrGlnGln SerLeuGlu AspTrpLeu GluAsnArgAsp

 424   GACGCCCGCA CCCGCAGCGT GCTGTACACC CAGTACATCG CCCTGGAGCT GGACTTCCTG
       AspAlaArg ThrArgSer ValLeuTyrThr GlnTyrIle AlaLeuGlu LeuAspPheLeu

 484   AACGCCATGC CCCTGTTCGC CATCCGCAAC CAGGAGGTGC CCCTGCTGAT GGTGTACGCC
       AsnAlaMet ProLeuPhe AlaIleArgAsn GlnGluVal ProLeuLeu MetValTyrAla

 544   CAGGCCGCCA ACCTGCACCT GCTGCTGCTG CGCGACGCCA GCCTGTTCGG CAGCGAGTTC
       GlnAlaAla AsnLeuHis LeuLeuLeuLeu ArgAspAla SerLeuPhe GlySerGluPhe

 604   GGCCTGACCA GCCAGGAGAT CCAGCGCTAC TACGAGCGCC AGGTGGAGCG CACCCGCGAC
       GlyLeuThr SerGlnGlu IleGlnArgTyr TyrGluArg GlnValGlu ArgThrArgAsp

 664   TACAGCGACT ACTGCGTGGA GTGGTACAAC ACCGGCCTGA ACAGCCTGCG CGGCACCAAC
       TyrSerAsp TyrCysVal GluTrpTyrAsn ThrGlyLeu AsnSerLeu ArgGlyThrAsn

 724   GCCGCCAGCT GGGTGCGCTA CAACCAGTTC CGCCGCGACC TGACCCTGGG CGTGCTGGAC
       AlaAlaSer TrpValArg TyrAsnGlnPhe ArgArgAsp LeuThrLeu GlyValLeuAsp

 784   CTGGTGGCCC TGTTCCCCAG CTACGACACC CGCACCTACC CCATCAACAC CAGCGCCCAG
       LeuValAla LeuPhePro SerTyrAspThr ArgThrTyr ProIleAsn ThrSerAlaGln

 844   CTGACCCGCG AGGTGTACAC CGACGCCATC GGCGCCACCG GCGTGAACAT GGCCAGCATG
       LeuThrArg GluValTyr ThrAspAlaIle GlyAlaThr GlyValAsn MetAlaSerMet

 904   AACTGGTACA ACAACAACGC CCCCAGCTTC AGCGCCATCG AGGCCGCCGC CATCCGCAGC
       AsnTrpTyr AsnAsnAsn AlaProSerPhe SerAlaIle GluAlaAla AlaIleArgSer

 964   CCCCACCTGC TGGACTTCCT GGAGCAGCTG ACCATCTTCA GCGCCAGCAG CCGCTGGAGC
       ProHisLeu LeuAspPhe LeuGluGlnLeu ThrIlePhe SerAlaSer SerArgTrpSer

1024   AACACCCGCC ACATGACCTA CTGGCGCGGC CACACCATCC AGAGCCGCCC CATCGGCGGC
       AsnThrArg HisMetThr TyrTrpArgGly HisThrIle GlnSerArg ProIleGlyGly
```

FIG. 6 (CONT)

```
1084   GGCCTGAACA CCAGCACCCA CGGCGCCACC AACACCAGCA TCAACCCCGT GACCCTGCGC
       GlyLeuAsn ThrSerThr HisGlyAlaThr AsnThrSer IleAsnPro ValThrLeuArg

1144   TTCGCCAGCC GCGACGTGTA CCGCACCGAG AGCTACGCCG GCGTGCTGCT GTGGGGCATC
       PheAlaSer ArgAspVal TyrArgThrGlu SerTyrAla GlyValLeu LeuTrpGlyIle

1204   TACCTGGAGC CCATCCACGG CGTGCCCACC GTGCGCTTCA ACTTCACCAA CCCCCAGAAC
       TyrLeuGlu ProIleHis GlyValProThr ValArgPhe AsnPheThr AsnProGlnAsn

1264   ATCAGCGACC GCGGCACCGC CAACTACAGC CAGCCCTACG AGAGCCCCGG CCTGCAGCTG
       IleSerAsp ArgGlyThr AlaAsnTyrSer GlnProTyr GluSerPro GlyLeuGlnLeu

1324   AAGGACAGCG AGACCGAGCT GCCCCCCGAG ACCACCGAGC GCCCCAACTA CGAGAGCTAC
       LysAspSer GluThrGlu LeuProProGlu ThrThrGlu ArgProAsn TyrGluSerTyr

1384   AGCCACCGCC TGAGCCACAT CGGCATCATC CTGCAGAGCC GCGTGAACGT GCCCGTGTAC
       SerHisArg LeuSerHis IleGlyIleIle LeuGlnSer ArgValAsn ValProValTyr

1444   AGCTGGACCC ACCGCAGCGC CGACCGCACC AACACCATCG GCCCCAACCG CATCACCCAG
       SerTrpThr HisArgSer AlaAspArgThr AsnThrIle GlyProAsn ArgIleThrGln

1504   ATCCCCATGG TGAAGGCCAG CGAGCTGCCC CAGGGCACCA CCGTGGTGCG CGGCCCCGGC
       IleProMet ValLysAla SerGluLeuPro GlnGlyThr ThrValVal ArgGlyProGly

1564   TTCACCGGCG GCGACATCCT GCGCCGCACC AACACCGGCG GCTTCGGCCC CATCCGCGTG
       PheThrGly GlyAspIle LeuArgArgThr AsnThrGly GlyPheGly ProIleArgVal

1624   ACCGTGAACG GCCCCCTGAC CCAGCGCTAC CGCATCGGCT TCCGCTACGC CAGCACCGTG
       ThrValAsn GlyProLeu ThrGlnArgTyr ArgIleGly PheArgTyr AlaSerThrVal

1684   GACTTCGACT TCTTCGTGAG CCGCGGCGGC ACCACCGTGA CAAACTTCCG CTTCCTGCGC
       AspPheAsp PhePheVal SerArgGlyGly ThrThrVal AsnAsnPhe ArgPheLeuArg

1744   ACCATGAACA GCGGCGACGA GCTGAAGTAC GGCAACTTCG TGCGCCGCGC CTTCACCACC
       ThrMetAsn SerGlyAsp GluLeuLysTyr GlyAsnPhe ValArgArg AlaPheThrThr

1804   CCCTTCACCT TCACCCAGAT CCAGGACATC ATCCGCACCA GCATCCAGGG CCTGAGCGGC
       ProPheThr PheThrGln IleGlnAspIle IleArgThr SerIleGln GlyLeuSerGly

1864   AACGGCGAGG TGTACATCGA CAAGATCGAG ATCATCCCCG TGACCGCCAC CTTCGAGGCC
       AsnGlyGlu ValTyrIle AspLysIleGlu IleIlePro ValThrAla ThrPheGluAla

1924   GAGTACGACC TGGAGCGCGC CCAGGAGGCC GTGAACGCCC TGTTCACCAA CACCAACCCC
       GluTyrAsp LeuGluArg AlaGlnGluAla ValAsnAla LeuPheThr AsnThrAsnPro

1984   CGCCGCCTGA AGACCGACGT GACCGACTAC CACATCGACC AGGTGAGCAA CCTGGTGGCC
       ArgArgLeu LysThrAsp ValThrAspTyr HisIleAsp GlnValSer AsnLeuValAla

2044   TGCCTGAGCG ACGAGTTCTG CCTGGACGAG AAGCGCGAGC TGCTGGAGAA GGTGAAGTAC
       CysLeuSer AspGluPhe CysLeuAspGlu LysArgGlu LeuLeuGlu LysValLysTyr
```

FIG. 6 (CONT)

```
2104  GCCAAGCGCC TGAGCGACGA GCGCAACCTG CTGCAGGACC CCAACTTCAC CAGCATCAAC
      AlaLysArg LeuSerAsp GluArgAsnLeu LeuGlnAsp ProAsnPhe ThrSerIleAsn

2164  AAGCAGCCCG ACTTCATCAG CACCAACGAG CAGAGCAACT TCACCAGCAT CCACGAGCAG
      LysGlnPro AspPheIle SerThrAsnGlu GlnSerAsn PheThrSer IleHisGluGln

2224  AGCGAGCACG GCTGGTGGGG CAGCGAGAAC ATCACCATCC AGGAGGGCAA CGACGTGTTC
      SerGluHis GlyTrpTrp GlySerGluAsn IleThrIle GlnGluGly AsnAspValPhe

2284  AAGGAGAACT ACGTGACCCT GCCCGGCACC TTCAACGAGT GCTACCCCAC CTACCTGTAC
      LysGluAsn TyrValThr LeuProGlyThr PheAsnGlu CysTyrPro ThrTyrLeuTyr

2344  CAGAAGATCG GCGAGAGCGA GCTGAAGGCC TACACCCGCT ACCAGCTGCG CGGCTACATC
      GlnLysIle GlyGluSer GluLeuLysAla TyrThrArg TyrGlnLeu ArgGlyTyrIle

2404  GAGGACAGCC AGGACCTGGA GATCTACCTG ATCCGCTACA ACGCCAAGCA CGAGACCCTG
      GluAspSer GlnAspLeu GluIleTyrLeu IleArgTyr AsnAlaLys HisGluThrLeu

2464  GACGTGCCCG GCACCGAGAG CCTGTGGCCC CTGAGCGTGG AGAGCCCCAT CGGCCGCTGC
      AspValPro GlyThrGlu SerLeuTrpPro LeuSerVal GluSerPro IleGlyArgCys

2524  GGCGAGCCCA ACCGCTGCGC CCCCCACTTC GAGTGGAACC CCGACCTGGA CTGCAGCTGC
      GlyGluPro AsnArgCys AlaProHisPhe GluTrpAsn ProAspLeu AspCysSerCys

2584  CGCGACGGCG AGAAGTGCGC CCACCACAGC CACCACTTCA GCCTGGACAT CGACGTGGGC
      ArgAspGly GluLysCys AlaHisHisSer HisHisPhe SerLeuAsp IleAspValGly

2644  TGCACCGACC TGCACGAGAA CCTGGGCGTG TGGGTGGTGT TCAAGATCAA GACCCAGGAG
      CysThrAsp LeuHisGlu AsnLeuGlyVal TrpValVal PheLysIle LysThrGlnGlu

2704  GGCCACGCCC GCCTGGGCAA CCTGGAGTTC ATCGAGGAGA AGCCCCTGCT GGGCGAGGCC
      GlyHisAla ArgLeuGly AsnLeuGluPhe IleGluGlu LysProLeu LeuGlyGluAla

2764  CTGAGCCGCG TGAAGCGCGC CGAGAAGAAG TGGCGCGACA AGCGCGAGAA GCTGCAGCTG
      LeuSerArg ValLysArg AlaGluLysLys TrpArgAsp LysArgGlu LysLeuGlnLeu

2824  GAGACCAAGC GCGTGTACAC CGAGGCCAAG GAGGCCGTGG ACGCCCTGTT CGTGGACAGC
      GluThrLys ArgValTyr ThrGluAlaLys GluAlaVal AspAlaLeu PheValAspSer

2884  CAGTACGACC GCCTGCAGGC CGACACCAAC ATCGGCATGA TCCACGCCGC CGACAAGCTG
      GlnTyrAsp ArgLeuGln AlaAspThrAsn IleGlyMet IleHisAla AlaAspLysLeu

2944  GTGCACCGCA TCCGCGAGGC CTACCTGAGC GAGCTGCCCG TGATCCCCGG CGTGAACGCC
      ValHisArg IleArgGlu AlaTyrLeuSer GluLeuPro ValIlePro GlyValAsnAla

3004  GAGATCTTCG AGGAGCTGGA GGGCCACATC ATCACCGCCA TCAGCCTGTA CGACGCCCGC
      GluIlePhe GluGluLeu GluGlyHisIle IleThrAla IleSerLeu TyrAspAlaArg
```

FIG. 6 (CONT)

```
3064   AACGTGGTGA AGAACGGCGA CTTCAACAAC GGCCTGACCT GCTGGAACGT GAAGGGCCAC
       AsnValVal LysAsnGly AspPheAsnAsn GlyLeuThr CysTrpAsn ValLysGlyHis

3124   GTGGACGTGC AGCAGAGCCA CCACCGCAGC GACCTGGTGA TCCCCGAGTG GGAGGCCGAG
       ValAspVal GlnGlnSer HisHisArgSer AspLeuVal IleProGlu TrpGluAlaGlu

3184   GTGAGCCAGG CCGTGCGCGT GTGCCCCGGC TGCGGCTACA TCCTGCGCGT GACCGCCTAC
       ValSerGln AlaValArg ValCysProGly CysGlyTyr IleLeuArg ValThrAlaTyr

3244   AAGGAGGGCT ACGGCGAGGG CTGCGTGACC ATCCACGAGA TCGAGAACAA CACCGACGAG
       LysGluGly TyrGlyGlu GlyCysValThr IleHisGlu IleGluAsn AsnThrAspGlu

3304   CTGAAGTTCA AGAACCGCGA GGAGGAGGAG GTGTACCCCA CCGACACCGG CACCTGCAAC
       LeuLysPhe LysAsnArg GluGluGluGlu ValTyrPro ThrAspThr GlyThrCysAsn

3364   GACTACACCG CCCACCAGGG CACCGCCGGC TGCGCCGACG CCTGCAACAG CCGCAACGCC
       AspTyrThr AlaHisGln GlyThrAlaGly CysAlaAsp AlaCysAsn SerArgAsnAla

3424   GGCTACGAGG ACGCCTACGA GGTGGACACC ACCGCCAGCG TGAACTACAA GCCCACCTAC
       GlyTyrGlu AspAlaTyr GluValAspThr ThrAlaSer ValAsnTyr LysProThrTyr

3484   GAGGAGGAGA CCTACACCGA CGTGCGCCGC GACAACCACT GCGAGTACGA CCGCGGCTAC
       GluGluGlu ThrTyrThr AspValArgArg AspAsnHis CysGluTyr AspArgGlyTyr

3544   GTGAACTACC CCCCCGTGCC CGCCGGCTAC GTGACCAAGG AGCTGGAGTA CTTCCCCGAG
       ValAsnTyr ProProVal ProAlaGlyTyr ValThrLys GluLeuGlu TyrPheProGlu

3604   ACCGACACCG TGTGGATCGA GATCGGCGAG ACCGAGGGCA AGTTCATCGT GGACAGCGTG
       ThrAspThr ValTrpIle GluIleGlyGlu ThrGluGly LysPheIle ValAspSerVal

3664   GAGCTGCTGC TGATGGAGGA GTAG
       GluLeuLeu LeuMetGlu Glu---
```

FIG. 7

SEQUENCE OF THE FULL-LENGTH HYBRID SYNTHETIC/NATIVE CRYIA(B) CHIMERIC GENE
The fusion point between the synthetic and native coding sequences is
indicated by a slash (/) in the sequence.

```
  1  ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGAG
     MetAspAsn AsnProAsn IleAsnGluCys IleProTyr AsnCysLeu SerAsnProGlu

 61  GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCCTG
     ValGluVal LeuGlyGly GluArgIleGlu ThrGlyTyr ThrProIle AspIleSerLeu

121  AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCCTG
     SerLeuThr GlnPheLeu LeuSerGluPhe ValProGly AlaGlyPhe ValLeuGlyLeu

181  GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGATC
     ValAspIle IleTrpGly IlePheGlyPro SerGlnTrp AspAlaPhe LeuValGlnIle

241  GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCCTG
     GluGlnLeu IleAsnGln ArgIleGluGlu PheAlaArg AsnGlnAla IleSerArgLeu

301  GAGGGCCTGA GCAACCTGTA CCAAATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGAC
     GluGlyLeu SerAsnLeu TyrGlnIleTyr AlaGluSer PheArgGlu TrpGluAlaAsp

361  CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGCC
     ProThrAsn ProAlaLeu ArgGluGluMet ArgIleGln PheAsnAsp MetAsnSerAla

421  CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGTG
     LeuThrThr AlaIlePro LeuPheAlaVal GlnAsnTyr GlnValPro LeuLeuSerVal

481  TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTCAGCGT GTTCGGCCAG
     TyrValGln AlaAlaAsn LeuHisLeuSer ValLeuArg AspValSer ValPheGlyGln

541  CGCTGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGATC
     ArgTrpGly PheAspAla AlaThrIleAsn SerArgTyr AsnAspLeu ThrArgLeuIle

601  GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG GCCTGGAGCG CGTGTGGGGT
     GlyAsnTyr ThrAspHis AlaValArgTrp TyrAsnThr GlyLeuGlu ArgValTrpGly

661  CCCGACAGCC GCGACTGGAT CAGGTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGTG
     ProAspSer ArgAspTrp IleArgTyrAsn GlnPheArg ArgGluLeu ThrLeuThrVal

721  CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGTG
     LeuAspIle ValSerLeu PheProAsnTyr AspSerArg ThrTyrPro IleArgThrVal

781  AGCCAGCTGA CCCGCGAGAT TTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC
     SerGlnLeu ThrArgGlu IleTyrThrAsn ProValLeu GluAsnPhe AspGlySerPhe

841  CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG
     ArgGlySer AlaGlnGly IleGluGlySer IleArgSer ProHisLeu MetAspIleLeu

901  AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG
     AsnSerIle ThrIleTyr ThrAspAlaHis ArgGlyGlu TyrTyrTrp SerGlyHisGln
```

FIG. 7 (CONT)

```
 961  ATCATGGCCA GCCCCGTCGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC
      IleMetAla SerProVal GlyPheSerGly ProGluPhe ThrPhePro LeuTyrGlyThr

1021  ATGGGCAACG CTGCACCTCA GCAGCGCATC GTGGCACAGC TGGGCCAGGG AGTGTACCGC
      MetGlyAsn AlaAlaPro GlnGlnArgIle ValAlaGln LeuGlyGln GlyValTyrArg

1081  ACCCTGAGCA GCACCCTGTA CCGTCGACCT TTCAACATCG GCATCAACAA CCAGCAGCTG
      ThrLeuSer SerThrLeu TyrArgArgPro PheAsnIle GlyIleAsn AsnGlnGlnLeu

1141  AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG
      SerValLeu AspGlyThr GluPheAlaTyr GlyThrSer SerAsnLeu ProSerAlaVal

1201  TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCTCAGAA CAACAACGTG
      TyrArgLys SerGlyThr ValAspSerLeu AspGluIle ProProGln AsnAsnAsnVal

1261  CCACCTCGAC AGGGCTTCAG CCACCGTCTG AGCCACGTGA GCATGTTCCG CAGTGGCTTC
      ProProArg GlnGlyPhe SerHisArgLeu SerHisVal SerMetPhe ArgSerGlyPhe

1321  AGCAACAGCA GCGTGAGCAT CATCCGTGCA CCTATGTTCA GCTGGATTCA CCGCAGTGCC
      SerAsnSer SerValSer IleIleArgAla ProMetPhe SerTrpIle HisArgSerAla

1381  GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC
      GluPheAsn AsnIleIle ProSerSerGln IleThrGln IleProLeu ThrLysSerThr

1441  AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG
      AsnLeuGly SerGlyThr SerValValLys GlyProGly PheThrGly GlyAspIleLeu

1501  CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC
      ArgArgThr SerProGly GlnIleSerThr LeuArgVal AsnIleThr AlaProLeuSer

1561  CAGCGCTACC GCGTCCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC
      GlnArgTyr ArgValArg IleArgTyrAla SerThrThr AsnLeuGln PheHisThrSer

1621  ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC
      IleAspGly ArgProIle AsnGlnGlyAsn PheSerAla ThrMetSer SerGlySerAsn

1681  CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC
      LeuGlnSer GlySerPhe ArgThrValGly PheThrThr ProPheAsn PheSerAsnGly

1741  AGCAGCGTGT TCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC
      SerSerVal PheThrLeu SerAlaHisVal PheAsnSer GlyAsnGlu ValTyrIleAsp

1801  CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCG AGTACGACCT GGAGAGGGCT
      ArgIleGlu PheValPro AlaGluValThr PheGluAla GluTyrAsp LeuGluArgAla

1861  CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG
      GlnLysAla ValAsnGlu LeuPheThrSer SerAsnGln IleGlyLeu LysThrAspVal
```

FIG. 7 (CONT)

1921 ACCGACTACC ACATCGAT/CA AGTATCCAAT TTAGTTGAGT GTTTATCTGATGAATTTTGT
      ThrAspTyr HisIleAsp/GlnValSerAsn LeuValGlu CysLeuSer AspGluPheCys

1981 CTGGATGAAA AAAAAGAATT GTCCGAGAAA GTCAAACATG CGAAGCGACT TAGTGATGAG
      LeuAspGlu LysLysGlu LeuSerGluLys ValLysHis AlaLysArg LeuSerAspGlu

2041 CGGAATTTAC TTCAAGATCC AAACTTTAGA GGGATCAATA GACAACTAGA CCGTGGCTGG
      ArgAsnLeu LeuGlnAsp ProAsnPheArg GlyIleAsn ArgGlnLeu AspArgGlyTrp

2101 AGAGGAAGTA CGGATATTAC CATCCAAGGA GGCGATGACG TATTCAAAGA GAATTACGTT
      ArgGlySer ThrAspIle ThrIleGlnGly GlyAspAsp ValPheLys GluAsnTyrVal

2161 ACGCTATTGG GTACCTTTGA TGAGTGCTAT CCAACGTATT TATATCAAAA AATAGATGAG
      ThrLeuLeu GlyThrPhe AspGluCysTyr ProThrTyr LeuTyrGln LysIleAspGlu

2221 TCGAAATTAA AAGCCTATAC CCGTTACCAA TTAAGAGGGT ATATCGAAGA TAGTCAAGAC
      SerLysLeu LysAlaTyr ThrArgTyrGln LeuArgGly TyrIleGlu AspSerGlnAsp

2281 TTAGAAATCT ATTTAATTCG CTACAATGCC AAACACGAAA CAGTAAATGT GCCAGGTACG
      LeuGluIle TyrLeuIle ArgTyrAsnAla LysHisGlu ThrValAsn ValProGlyThr

2341 GGTTCCTTAT GGCCGCTTTC AGCCCCAAGT CCAATCGGAA AATGTGCCCA TCATTCCCAT
      GlySerLeu TrpProLeu SerAlaProSer ProIleGly LysCysAla HisHisSerHis

2401 CATTTCTCCT TGGACATTGA TGTTGGATGT ACAGACTTAA ATGAGGACTT AGGTGTATGG
      HisPheSer LeuAspIle AspValGlyCys ThrAspLeu AsnGluAsp LeuGlyValTrp

2461 GTGATATTCA AGATTAAGAC GCAAGATGGC CATGCAAGAC TAGGAAATCT AGAATTTCTC
      ValIlePhe LysIleLys ThrGlnAspGly HisAlaArg LeuGlyAsn LeuGluPheLeu

2521 GAAGAGAAAC CATTAGTAGG AGAAGCACTA GCTCGTGTGA AAAGAGCGGA GAAAAAATGG
      GluGluLys ProLeuVal GlyGluAlaLeu AlaArgVal LysArgAla GluLysLysTrp

2581 AGAGACAAAC GTGAAAAATT GGAATGGGAA ACAAATATTG TTTATAAAGA GGCAAAAGAA
      ArgAspLys ArgGluLys LeuGluTrpGlu ThrAsnIle ValTyrLys GluAlaLysGlu

2641 TCTGTAGATG CTTTATTTGT AAACTCTCAA TATGATAGAT TACAAGCGGA TACCAACATC
      SerValAsp AlaLeuPhe ValAsnSerGln TyrAspArg LeuGlnAla AspThrAsnIle

2701 GCGATGATTC ATGCGGCAGA TAAACGCGTT CATAGCATTC GAGAAGCTTA TCTGCCTGAG
      AlaMetIle HisAlaAla AspLysArgVal HisSerIle ArgGluAla TyrLeuProGlu

2761 CTGTCTGTGA TTCCGGGTGT CAATGCGGCT ATTTTTGAAG AATTAGAAGG GCGTATTTTC
      LeuSerVal IleProGly ValAsnAlaAla IlePheGlu GluLeuGlu GlyArgIlePhe

2821 ACTGCATTCT CCCTATATGA TGCGAGAAAT GTCATTAAAA ATGGTGATTT TAATAATGGC
      ThrAlaPhe SerLeuTyr AspAlaArgAsn ValIleLys AsnGlyAsp PheAsnAsnGly

FIG. 7 (CONT)

2881   TTATCCTGCT GGAACGTGAA AGGGCATGTA GATGTAGAAG AACAAAACAA CCACCGTTCG
       LeuSerCys TrpAsnVal LysGlyHisVal AspValGlu GluGlnAsn AsnHisArgSer

2941   GTCCTTGTTG TTCCGGAATG GGAAGCAGAA GTGTCACAAG AAGTTCGTGT CTGTCCGGGT
       ValLeuVal ValProGlu TrpGluAlaGlu ValSerGln GluValArg ValCysProGly

3001   CGTGGCTATA TCCTTCGTGT CACAGCGTAC AAGGAGGGAT ATGGAGAAGG TTGCGTAACC
       ArgGlyTyr IleLeuArg ValThrAlaTyr LysGluGly TyrGlyGlu GlyCysValThr

3061   ATTCATGAGA TCGAGAACAA TACAGACGAA CTGAAGTTTA GCAACTGTGT AGAAGAGGAA
       IleHisGlu IleGluAsn AsnThrAspGlu LeuLysPhe SerAsnCys ValGluGluGlu

3121   GTATATCCAA ACAACACGGT AACGTGTAAT GATTATACTG CGACTCAAGA AGAATATGAG
       ValTyrPro AsnAsnThr ValThrCysAsn AspTyrThr AlaThrGln GluGluTyrGlu

3181   GGTACGTACA CTTCTCGTAA TCGAGGATAT GACGGAGCCT ATGAAAGCAA TTCTTCTGTA
       GlyThrTyr ThrSerArg AsnArgGlyTyr AspGlyAla TyrGluSer AsnSerSerVal

3241   CCAGCTGATT ATGCATCAGC CTATGAAGAA AAAGCATATA CAGATGGACG AAGAGACAAT
       ProAlaAsp TyrAlaSer AlaTyrGluGlu LysAlaTyr ThrAspGly ArgArgAspAsn

3301   CCTTGTGAAT CTAACAGAGG ATATGGGGAT TACACACCAC TACCAGCTGG CTATGTGACA
       ProCysGlu SerAsnArg GlyTyrGlyAsp TyrThrPro LeuProAla GlyTyrValThr

3361   AAAGAATTAG AGTACTTCCC AGAAACCGAT AAGGTATGGA TTGAGATCGG AGAAACGGAA
       LysGluLeu GluTyrPhe ProGluThrAsp LysValTrp IleGluIle GlyGluThrGlu

3421   GGAACATTCA TCGTGGACAG CGTGGAATTA CTTCTTATGG AGGAATAA
       GlyThrPhe IleValAsp SerValGluLeu LeuLeuMet GluGlu---

FIG. 8

FIG. 9

```
  1  ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGAG
     MetAspAsn AsnProAsn IleAsnGluCys IleProTyr AsnCysLeu SerAsnProGlu

 61  GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCCTG
     ValGluVal LeuGlyGly GluArgIleGlu ThrGlyTyr ThrProIle AspIleSerLeu

121  AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCCTG
     SerLeuThr GlnPheLeu LeuSerGluPhe ValProGly AlaGlyPhe ValLeuGlyLeu

181  GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGATC
     ValAspIle IleTrpGly IlePheGlyPro SerGlnTrp AspAlaPhe LeuValGlnIle

241  GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCCTG
     GluGlnLeu IleAsnGln ArgIleGluGlu PheAlaArg AsnGlnAla IleSerArgLeu

301  GAGGGCCTGA GCAACCTGTA CCAAATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGAC
     GluGlyLeu SerAsnLeu TyrGlnIleTyr AlaGluSer PheArgGlu TrpGluAlaAsp

361  CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGCC
     ProThrAsn ProAlaLeu ArgGluGluMet ArgIleGln PheAsnAsp MetAsnSerAla

421  CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGTG
     LeuThrThr AlaIlePro LeuPheAlaVal GlnAsnTyr GlnValPro LeuLeuSerVal

481  TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTCAGCGT GTTCGGCCAG
     TyrValGln AlaAlaAsn LeuHisLeuSer ValLeuArg AspValSer ValPheGlyGln

541  CGCTGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGATC
     ArgTrpGly PheAspAla AlaThrIleAsn SerArgTyr AsnAspLeu ThrArgLeuIle

601  GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG GCCTGGAGCG CGTGTGGGGT
     GlyAsnTyr ThrAspHis AlaValArgTrp TyrAsnThr GlyLeuGlu ArgValTrpGly

661  CCCGACAGCC GCGACTGGAT CAGGTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGTG
     ProAspSer ArgAspTrp IleArgTyrAsn GlnPheArg ArgGluLeu ThrLeuThrVal

721  CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGTG
     LeuAspIle ValSerLeu PheProAsnTyr AspSerArg ThrTyrPro IleArgThrVal

781  AGCCAGCTGA CCCGCGAGAT TTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC
     SerGlnLeu ThrArgGlu IleTyrThrAsn ProValLeu GluAsnPhe AspGlySerPhe

841  CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG
     ArgGlySer AlaGlnGly IleGluGlySer IleArgSer ProHisLeu MetAspIleLeu

901  AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG
     AsnSerIle ThrIleTyr ThrAspAlaHis ArgGlyGlu TyrTyrTrp SerGlyHisGln

961  ATCATGGCCA GCCCCGTCGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC
     IleMetAla SerProVal GlyPheSerGly ProGluPhe ThrPhePro LeuTyrGlyThr
```

FIG. 9 (CONT)

```
1021   ATGGGCAACG CTGCACCTCA GCAGCGCATC GTGGCACAGC TGGGCCAGGG AGTGTACCGC
       MetGlyAsn AlaAlaPro GlnGlnArgIle ValAlaGln LeuGlyGln GlyValTyrArg

1081   ACCCTGAGCA GCACCCTGTA CCGTCGACCT TTCAACATCG GCATCAACAA CCAGCAGCTG
       ThrLeuSer SerThrLeu TyrArgArgPro PheAsnIle GlyIleAsn AsnGlnGlnLeu

1141   AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG
       SerValLeu AspGlyThr GluPheAlaTyr GlyThrSer SerAsnLeu ProSerAlaVal

1201   TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCTCAGAA CAACAACGTG
       TyrArgLys SerGlyThr ValAspSerLeu AspGluIle ProProGln AsnAsnAsnVal

1261   CCACCTCGAC AGGGCTTCAG CCACCGTCTG AGCCACGTGA GCATGTTCCG CAGTGGCTTC
       ProProArg GlnGlyPhe·SerHisArgLeu SerHisVal SerMetPhe ArgSerGlyPhe

1321   AGCAACAGCA GCGTGAGCAT CATCCGTGCA CCTATGTTCA GCTGGATTCA CCGCAGTGCC
       SerAsnSer SerValSer IleIleArgAla ProMetPhe SerTrpIle HisArgSerAla

1381   GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC
       GluPheAsn AsnIleIle ProSerSerGln IleThrGln IleProLeu ThrLysSerThr

1441   AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG
       AsnLeuGly SerGlyThr SerValValLys GlyProGly PheThrGly GlyAspIleLeu

1501   CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC
       ArgArgThr SerProGly GlnIleSerThr LeuArgVal AsnIleThr AlaProLeuSer

1561   CAGCGCTACC GCGTCCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC
       GlnArgTyr ArgValArg IleArgTyrAla SerThrThr AsnLeuGln PheHisThrSer

1621   ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC
       IleAspGly ArgProIle AsnGlnGlyAsn PheSerAla ThrMetSer SerGlySerAsn

1681   CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC
       LeuGlnSer GlySerPhe ArgThrValGly PheThrThr ProPheAsn PheSerAsnGly

1741   AGCAGCGTGT TCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC
       SerSerVal PheThrLeu SerAlaHisVal PheAsnSer GlyAsnGlu ValTyrIleAsp

1801   CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCG AGTACGACCT GGAGAGGGCT
       ArgIleGlu PheValPro AlaGluValThr PheGluAla GluTyrAsp LeuGluArgAla

1861   CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG
       GlnLysAla ValAsnGlu LeuPheThrSer SerAsnGln IleGlyLeu LysThrAspVal

1921   ACCGACTACC ACATCGATCA GTATCCAAT TTAGTTGAGT GTTTATCTGA TGAATTTTGT
       ThrAspTyr HisIleAsp GlnValSerAsn LeuValGlu CysLeuSer AspGluPheCys

1981   CTGGATGAAA AAAAGAATT GTCCGAGAAA GTCAAACATG CGAAGCGACT TAGTGATGAG
       LeuAspGlu LysLysGlu LeuSerGluLys ValLysHis AlaLysArg LeuSerAspGlu
```

FIG. 9 (CONT)

2041 CGGAATTTAC TTCAAGATCC AAACTTTAGA GGGATCAATA GACAACTAGA CCGTGGCTGG
      ArgAsnLeu LeuGlnAsp ProAsnPheArg GlyIleAsn ArgGlnLeu AspArgGlyTrp

2101 AGAGGAAGTA CGGATATTAC CATCCAAGGA GGCGATGACG TATTCAAAGA GAATTACGTT
      ArgGlySer ThrAspIle ThrIleGlnGly GlyAspAsp ValPheLys GluAsnTyrVal

2161 ACGCTATTGG GTACCTTCGA CGAGTGCTAC CCCACCTACC TGTACCAGAA GATCGACGAG
      ThrLeuLeu GlyThrPhe AspGluCysTyr ProThrTyr LeuTyrGln LysIleAspGlu

2221 AGCAAGCTGA AGGCCTACAC CCGCTACCAG CTGCGCGGCT ACATCGAGGA CAGCCAGGAC
      SerLysLeu LysAlaTyr ThrArgTyrGln LeuArgGly TyrIleGlu AspSerGlnAsp

2281 CTGGAAATCT ACCTGATCCG CTACAACGCC AAGCACGAGA CCGTGAACGT GCCCGGCACC
      LeuGluIle TyrLeuIle ArgTyrAsnAla LysHisGlu ThrValAsn ValProGlyThr

2341 GGCAGCCTGT GGCCCCTGAG CGCCCCCAGC CCCATCGGCA AGTGCGGGGA GCCGAATCGA
      GlySerLeu TrpProLeu SerAlaProSer ProIleGly LysCysGly GluProAsnArg

2401 TGCGCTCCGC ACCTGGAGTG GAACCCGGAC CTAGACTGCA GCTGCAGGGA CGGGGAGAAG
      CysAlaPro HisLeuGlu TrpAsnProAsp LeuAspCys SerCysArg AspGlyGluLys

2461 TGCGCCCACC ACAGCCACCA CTTCAGCCTG GACATCGACG TGGGCTGCAC CGACCTGAAC
      CysAlaHis HisSerHis HisPheSerLeu AspIleAsp ValGlyCys ThrAspLeuAsn

2521 GAGGACCTGG GCGTGTGGGT GATCTTCAAG ATCAAGACCC AGGACGGCCA CGCCCGCCTG
      GluAspLeu GlyValTrp ValIlePheLys IleLysThr GlnAspGly HisAlaArgLeu

2581 GGCAATCTAG AATTTCTCGA AGAGAAACCA TTAGTAGGAG AAGCACTAGC TCGTGTGAAA
      GlyAsnLeu GluPheLeu GluGluLysPro LeuValGly GluAlaLeu AlaArgValLys

2641 AGAGCGGAGA AAAAATGGAG AGACAAACGT GAAAAATTGG AATGGGAAAC AAATATTGTT
      ArgAlaGlu LysLysTrp ArgAspLysArg GluLysLeu GluTrpGlu ThrAsnIleVal

2701 TATAAAGAGG CAAAAGAATC TGTAGATGCT TTATTTGTAA ACTCTCAATA TGATAGATTA
      TyrLysGlu AlaLysGlu SerValAspAla LeuPheVal AsnSerGln TyrAspArgLeu

2761 CAAGCGGATA CCAACATCGC GATGATTCAT GCGGCAGATA AACGCGTTCA TAGCATTCGA
      GlnAlaAsp ThrAsnIle AlaMetIleHis AlaAlaAsp LysArgVal HisSerIleArg

2821 GAAGCTTATC TGCCTGAGCT GTCTGTGATT CCGGGTGTCA ATGCGGCTAT TTTTGAAGAA
      GluAlaTyr LeuProGlu LeuSerValIle ProGlyVal AsnAlaAla IlePheGluGlu

2881 TTAGAAGGGC GTATTTTCAC TGCATTCTCC CTATATGATG CGAGAAATGT CATTAAAAAT
      LeuGluGly ArgIlePhe ThrAlaPheSer LeuTyrAsp AlaArgAsn ValIleLysAsn

2941 GGTGATTTTA ATAATGGCTT ATCCTGCTGG AACGTGAAAG GGCATGTAGA TGTAGAAGAA
      GlyAspPhe AsnAsnGly LeuSerCysTrp AsnValLys GlyHisVal AspValGluGlu

3001 CAAAACAACC ACCGTTCGGT CCTTGTTGTT CCGGAATGGG AAGCAGAAGT GTCACAAGAA
      GlnAsnAsn HisArgSer ValLeuValVal ProGluTrp GluAlaGlu ValSerGlnGlu

38/94

FIG. 9 (CONT)

```
3061   GTTCGTGTCT GTCCGGGTCG TGGCTATATC CTTCGTGTCA CAGCGTACAA GGAGGGATAT
       ValArgVal CysProGly ArgGlyTyrIle LeuArgVal ThrAlaTyr LysGluGlyTyr

3121   GGAGAAGGTT GCGTAACCAT TCATGAGATC GAGAACAATA CAGACGAACT GAAGTTTAGC
       GlyGluGly CysValThr IleHisGluIle GluAsnAsn ThrAspGlu LeuLysPheSer

3181   AACTGTGTAG AAGAGGAAGT ATATCCAAAC AACACGGTAA CGTGTAATGA TTATACTGCG
       AsnCysVal GluGluGlu ValTyrProAsn AsnThrVal ThrCysAsn AspTyrThrAla

3241   ACTCAAGAAG AATATGAGGG TACGTACACT TCTCGTAATC GAGGATATGA CGGAGCCTAT
       ThrGlnGlu GluTyrGlu GlyThrTyrThr SerArgAsn ArgGlyTyr AspGlyAlaTyr

3301   GAAAGCAATT CTTCTGTACC AGCTGATTAT GCATCAGCCT ATGAAGAAAA AGCATATACA
       GluSerAsn SerSerVal ProAlaAspTyr AlaSerAla TyrGluGlu LysAlaTyrThr

3361   GATGGACGAA GAGACAATCC TTGTGAATCT AACAGAGGAT ATGGGGATTA CACACCACTA
       AspGlyArg ArgAspAsn ProCysGluSer AsnArgGly TyrGlyAsp TyrThrProLeu

3421   CCAGCTGGCT ATGTGACAAA AGAATTAGAG TACTTCCCAG AAACCGATAA GGTATGGATT
       ProAlaGly TyrValThr LysGluLeuGlu TyrPhePro GluThrAsp LysValTrpIle

3481   GAGATCGGAG AAACGGAAGG AACATTCATC GTGGACAGCG TGGAATTACT TCTTATGGAG
       GluIleGly GluThrGlu GlyThrPheIle ValAspSer ValGluLeu LeuLeuMetGlu

3541   GAATAA
       Glu---
```

FIG. 10

FIG. 11

```
  1  ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGA:
     MetAspAsn AsnProAsn IleAsnGluCys IleProTyr AsnCysLeu SerAsnProG::

 61  GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCC::
     ValGluVal LeuGlyGly GluArgIleGlu ThrGlyTyr ThrProIle AspIleSerLe:

121  AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCC::
     SerLeuThr GlnPheLeu LeuSerGluPhe ValProGly AlaGlyPhe ValLeuGlyLe:

181  GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGA::
     ValAspIle IleTrpGly IlePheGlyPro SerGlnTrp AspAlaPhe LeuValGlnI:e

241  GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCC::
     GluGlnLeu IleAsnGln ArgIleGluGlu PheAlaArg AsnGlnAla IleSerArgLe:

301  GAGGGCCTGA GCAACCTGTA CCAAATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGA:
     GluGlyLeu SerAsnLeu TyrGlnIleTyr AlaGluSer PheArgGlu TrpGluAlaAs:

361  CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGC:
     ProThrAsn ProAlaLeu ArgGluGluMet ArgIleGln PheAsnAsp MetAsnSerA:a

421  CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGT:
     LeuThrThr AlaIlePro LeuPheAlaVal GlnAsnTyr GlnValPro LeuLeuSerVa:

481  TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTCAGCGT GTTCGGCCA:
     TyrValGln AlaAlaAsn LeuHisLeuSer ValLeuArg AspValSer ValPheGlyG::

541  CGCTGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGA::
     ArgTrpGly PheAspAla AlaThrIleAsn SerArgTyr AsnAspLeu ThrArgLeuI:e

601  GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG GCCTGGAGCG CGTGTGGGG:
     GlyAsnTyr ThrAspHis AlaValArgTrp TyrAsnThr GlyLeuGlu ArgValTrpGl;

661  CCCGACAGCC GCGACTGGAT CAGGTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGT:
     ProAspSer ArgAspTrp IleArgTyrAsn GlnPheArg ArgGluLeu ThrLeuThrVa:

721  CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGT:
     LeuAspIle ValSerLeu PheProAsnTyr AspSerArg ThrTyrPro IleArgThrVa:

781  AGCCAGCTGA CCCGCGAGAT TTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC
     SerGlnLeu ThrArgGlu IleTyrThrAsn ProValLeu GluAsnPhe AspGlySerPhe

841  CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG
     ArgGlySer AlaGlnGly IleGluGlySer IleArgSer ProHisLeu MetAspIleLeu

901  AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG
     AsnSerIle ThrIleTyr ThrAspAlaHis ArgGlyGlu TyrTyrTrp SerGlyHisGln

961  ATCATGGCCA GCCCCGTCGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC
     IleMetAla SerProVal GlyPheSerGly ProGluPhe ThrPhePro LeuTyrGlyThr
```

FIG. 11 (CONT)

1021   ATGGGCAACG CTGCACCTCA GCAGCGCATC GTGGCACAGC TGGGCCAGGG AGTGTACCGC
       MetGlyAsn AlaAlaPro GlnGlnArgIle ValAlaGln LeuGlyGln GlyValTyrArg

1081   ACCCTGAGCA GCACCCTGTA CCGTCGACCT TTCAACATCG GCATCAACAA CCAGCAGCTG
       ThrLeuSer SerThrLeu TyrArgArgPro PheAsnIle GlyIleAsn AsnGlnGlnLeu

1141   AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG
       SerValLeu AspGlyThr GluPheAlaTyr GlyThrSer SerAsnLeu ProSerAlaVal

1201   TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCTCAGAA CAACAACGTG
       TyrArgLys SerGlyThr ValAspSerLeu AspGluIle ProProGln AsnAsnAsnVal

1261   CCACCTCGAC AGGGCTTCAG CCACCGTCTG AGCCACGTGA GCATGTTCCG CAGTGGCTTC
       ProProArg GlnGlyPhe SerHisArgLeu SerHisVal SerMetPhe ArgSerGlyPhe

1321   AGCAACAGCA GCGTGAGCAT CATCCGTGCA CCTATGTTCA GCTGGATTCA CCGCAGTGCC
       SerAsnSer SerValSer IleIleArgAla ProMetPhe SerTrpIle HisArgSerAla

1381   GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC
       GluPheAsn AsnIleIle ProSerSerGln IleThrGln IleProLeu ThrLysSerThr

1441   AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG
       AsnLeuGly SerGlyThr SerValValLys GlyProGly PheThrGly GlyAspIleLeu

1501   CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC
       ArgArgThr SerProGly GlnIleSerThr LeuArgVal AsnIleThr AlaProLeuSer

1561   CAGCGCTACC GCGTCCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC
       GlnArgTyr ArgValArg IleArgTyrAla SerThrThr AsnLeuGln PheHisThrSer

1621   ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC
       IleAspGly ArgProIle AsnGlnGlyAsn PheSerAla ThrMetSer SerGlySerAsn

1681   CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC
       LeuGlnSer GlySerPhe ArgThrValGly PheThrThr ProPheAsn PheSerAsnGly

1741   AGCAGCGTGT TCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC
       SerSerVal PheThrLeu SerAlaHisVal PheAsnSer GlyAsnGlu ValTyrIleAsp

1801   CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCG AGTACGACCT GGAGAGGGCT
       ArgIleGlu PheValPro AlaGluValThr PheGluAla GluTyrAsp LeuGluArgAla

1861   CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG
       GlnLysAla ValAsnGlu LeuPheThrSer SerAsnGln IleGlyLeu LysThrAspVal

1921   ACCGACTACC ACATCGATCA GGTGAGCAAC CTGGTGGAGT GCTTAAGCGA CGAGTTCTGC
       ThrAspTyr HisIleAsp GlnValSerAsn LeuValGlu CysLeuSer AspGluPheCys

1981   CTGGACGAGA GAAGGAGCT GAGCGAGAAG GTGAAGCACG CCAAGCGCCT GAGCGACGAG
       LeuAspGlu LysLysGlu LeuSerGluLys ValLysHis AlaLysArg LeuSerAspGlu

FIG. 11 (CONT)

```
2041  CGCAACCTGC TGCAGGACCC CAACTTCCGC GGCATCAACC GCCAGCTGGA CCGCGGCTGG
      ArgAsnLeu LeuGlnAsp ProAsnPheArg GlyIleAsn ArgGlnLeu AspArgGlyTrp

2101  CGAGGCAGCA CCGATATCAC CATCCAGGGC GGCGACGACG TGTTCAAGGA GAACTACGTG
      ArgGlySer ThrAspIle ThrIleGlnGly GlyAspAsp ValPheLys GluAsnTyrVal

2161  ACCCTGCTGG GCACCTTCGA CGAGTGCTAC CCCACCTACC TGTACCAGAA GATCGACGAG
      ThrLeuLeu GlyThrPhe AspGluCysTyr ProThrTyr LeuTyrGln LysIleAspGlu

2221  AGCAAGCTGA AGGCCTACAC CCGCTACCAG CTGCGCGGCT ACATCGAGGA CAGCCAGGAC
      SerLysLeu LysAlaTyr ThrArgTyrGln LeuArgGly TyrIleGlu AspSerGlnAsp

2281  CTGGAAATCT ACCTGATCCG CTACAACGCC AAGCACGAGA CCGTGAACGT GCCCGGCACC
      LeuGluIle TyrLeuIle ArgTyrAsnAla LysHisGlu ThrValAsn ValProGlyThr

2341  GGCAGCCTGT GGCCCCTGAG CGCCCCCAGC CCCATCGGCA AGTGCGGGGA GCCGAATCGA
      GlySerLeu TrpProLeu SerAlaProSer ProIleGly LysCysGly GluProAsnArg

2401  TGCGCTCCGC ACCTGGAGTG GAACCCGGAC CTAGACTGCA GCTGCAGGGA CGGGGAGAAG
      CysAlaPro HisLeuGlu TrpAsnProAsp LeuAspCys SerCysArg AspGlyGluLys

2461  TGCGCCCACC ACAGCCACCA CTTCAGCCTG GACATCGACG TGGGCTGCAC CGACCTGAAC
      CysAlaHis HisSerHis HisPheSerLeu AspIleAsp ValGlyCys ThrAspLeuAsn

2521  GAGGACCTGG GCGTGTGGGT GATCTTCAAG ATCAAGACCC AGGACGGCCA CGCCCGCCTG
      GluAspLeu GlyValTrp ValIlePheLys IleLysThr GlnAspGly HisAlaArgLeu

2581  GGCAATCTAG AATTTCTCGA AGAGAAACCA TTAGTAGGAG AAGCACTAGC TCGTGTGAAA
      GlyAsnLeu GluPheLeu GluGluLysPro LeuValGly GluAlaLeu AlaArgValLys

2641  AGAGCGGAGA AAAAATGGAG AGACAAACGT GAAAAATTGG AATGGGAAAC AAATATTGTT
      ArgAlaGlu LysLysTrp ArgAspLysArg GluLysLeu GluTrpGlu ThrAsnIleVal

2701  TATAAAGAGG CAAAAGAATC TGTAGATGCT TTATTTGTAA ACTCTCAATA TGATAGATTA
      TyrLysGlu AlaLysGlu SerValAspAla LeuPheVal AsnSerGln TyrAspArgLeu

2761  CAAGCGGATA CCAACATCGC GATGATTCAT GCGGCAGATA AACGCGTTCA TAGCATTCGA
      GlnAlaAsp ThrAsnIle AlaMetIleHis AlaAlaAsp LysArgVal HisSerIleArg

2821  GAAGCTTATC TGCCTGAGCT GTCTGTGATT CCGGGTGTCA ATGCGGCTAT TTTTGAAGAA
      GluAlaTyr LeuProGlu LeuSerValIle ProGlyVal AsnAlaAla IlePheGluGlu

2881  TTAGAAGGGC GTATTTTCAC TGCATTCTCC CTATATGATG CGAGAAATGT CATTAAAAAT
      LeuGluGly ArgIlePhe ThrAlaPheSer LeuTyrAsp AlaArgAsn ValIleLysAsn

2941  GGTGATTTTA ATAATGGCTT ATCCTGCTGG AACGTGAAAG GGCATGTAGA TGTAGAAGAA
      GlyAspPhe AsnAsnGly LeuSerCysTrp AsnValLys GlyHisVal AspValGluGlu

3001  CAAAACAACC ACCGTTCGGT CCTTGTTGTT CCGGAATGGG AAGCAGAAGT GTCACAAGAA
      GlnAsnAsn HisArgSer ValLeuValVal ProGluTrp GluAlaGlu ValSerGlnGlu
```

FIG. 11 (CONT)

```
3061  GTTCGTGTCT GTCCGGGTCG TGGCTATATC CTTCGTGTCA CAGCGTACAA GGAGGGATAT
      ValArgVal CysProGly ArgGlyTyrIle LeuArgVal ThrAlaTyr LysGluGlyTyr

3121  GGAGAAGGTT GCGTAACCAT TCATGAGATC GAGAACAATA CAGACGAACT GAAGTTTAGC
      GlyGluGly CysValThr IleHisGluIle GluAsnAsn ThrAspGlu LeuLysPheSer

3181  AACTGTGTAG AAGAGGAAGT ATATCCAAAC AACACGGTAA CGTGTAATGA TTATACTGCG
      AsnCysVal GluGluGlu ValTyrProAsn AsnThrVal ThrCysAsn AspTyrThrAla

3241  ACTCAAGAAG AATATGAGGG TACGTACACT TCTCGTAATC GAGGATATGA CGGAGCCTAT
      ThrGlnGlu GluTyrGlu GlyThrTyrThr SerArgAsn ArgGlyTyr AspGlyAlaTyr

3301  GAAAGCAATT CTTCTGTACC AGCTGATTAT GCATCAGCCT ATGAAGAAAA AGCATATACA
      GluSerAsn SerSerVal ProAlaAspTyr AlaSerAla TyrGluGlu LysAlaTyrThr

3361  GATGGACGAA GAGACAATCC TTGTGAATCT AACAGAGGAT ATGGGGATTA CACACCACTA
      AspGlyArg ArgAspAsn ProCysGluSer AsnArgGly TyrGlyAsp TyrThrProLeu

3421  CCAGCTGGCT ATGTGACAAA AGAATTAGAG TACTTCCCAG AAACCGATAA GGTATGGATT
      ProAlaGly TyrValThr LysGluLeuGlu TyrPhePro GluThrAsp LysValTrpIle

3481  GAGATCGGAG AAACGGAAGG AACATTCATC GTGGACAGCG TGGAATTACT TCTTATGGAG
      GluIleGly GluThrGlu GlyThrPheIle ValAspSer ValGluLeu LeuLeuMetGlu

3541  GAATAA
      Glu---
```

FIG. 12

FIG. 13

```
  1  ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGAG
     MetAspAsn  AsnProAsn  IleAsnGluCys IleProTyr  AsnCysLeu  SerAsnProGlu

 61  GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCCTG
     ValGluVal  LeuGlyGly  GluArgIleGlu ThrGlyTyr  ThrProIle  AspIleSerLeu

121  AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCCTG
     SerLeuThr  GlnPheLeu  LeuSerGluPhe ValProGly  AlaGlyPhe  ValLeuGlyLeu

181  GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGATC
     ValAspIle  IleTrpGly  IlePheGlyPro SerGlnTrp  AspAlaPhe  LeuValGlnIle

241  GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCCTG
     GluGlnLeu  IleAsnGln  ArgIleGluGlu PheAlaArg  AsnGlnAla  IleSerArgLeu

301  GAGGGCCTGA GCAACCTGTA CCAAATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGAC
     GluGlyLeu  SerAsnLeu  TyrGlnIleTyr AlaGluSer  PheArgGlu  TrpGluAlaAsp

361  CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGCC
     ProThrAsn  ProAlaLeu  ArgGluGluMet ArgIleGln  PheAsnAsp  MetAsnSerAla

421  CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGTG
     LeuThrThr  AlaIlePro  LeuPheAlaVal GlnAsnTyr  GlnValPro  LeuLeuSerVal

481  TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTCAGCGT GTTCGGCCAG
     TyrValGln  AlaAlaAsn  LeuHisLeuSer ValLeuArg  AspValSer  ValPheGlyGln

541  CGCTGGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGATC
     ArgTrpGly  PheAspAla  AlaThrIleAsn SerArgTyr  AsnAspLeu  ThrArgLeuIle

601  GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG CCTGGAGCG CGTGTGGGGT
     GlyAsnTyr  ThrAspHis  AlaValArgTrp TyrAsnThr  GlyLeuGlu  ArgValTrpGly

661  CCCGACAGCC GCGACTGGAT CAGGTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGTG
     ProAspSer  ArgAspTrp  IleArgTyrAsn GlnPheArg  ArgGluLeu  ThrLeuThrVal

721  CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGTG
     LeuAspIle  ValSerLeu  PheProAsnTyr AspSerArg  ThrTyrPro  IleArgThrVal

781  AGCCAGCTGA CCCGCGAGAT TTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC
     SerGlnLeu  ThrArgGlu  IleTyrThrAsn ProValLeu  GluAsnPhe  AspGlySerPhe

841  CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG
     ArgGlySer  AlaGlnGly  IleGluGlySer IleArgSer  ProHisLeu  MetAspIleLeu

901  AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG
     AsnSerIle  ThrIleTyr  ThrAspAlaHis ArgGlyGlu  TyrTyrTrp  SerGlyHisGln

961  ATCATGGCCA GCCCCGTCGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC
     IleMetAla  SerProVal  GlyPheSerGly ProGluPhe  ThrPhePro  LeuTyrGlyThr
```

46/94

FIG. 13 (CONT)

1021 ATGGGCAACG CTGCACCTCA GCAGCGCATC GTGGCACAGC TGGGCCAGGG AGTGTACCGC
      MetGlyAsn AlaAlaPro GlnGlnArgIle ValAlaGln LeuGlyGln GlyValTyrArg

1081 ACCCTGAGCA GCACCCTGTA CCGTCGACCT TTCAACATCG GCATCAACAA CCAGCAGCTG
      ThrLeuSer SerThrLeu TyrArgArgPro PheAsnIle GlyIleAsn AsnGlnGlnLeu

1141 AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG
      SerValLeu AspGlyThr GluPheAlaTyr GlyThrSer SerAsnLeu ProSerAlaVal

1201 TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCTCAGAA CAACAACGTG
      TyrArgLys SerGlyThr ValAspSerLeu AspGluIle ProProGln AsnAsnAsnVal

1261 CCACCTCGAC AGGGCTTCAG CCACCGTCTG AGCCACGTGA GCATGTTCCG CAGTGGCTTC
      ProProArg GlnGlyPhe SerHisArgLeu SerHisVal SerMetPhe ArgSerGlyPhe

1321 AGCAACAGCA GCGTGAGCAT CATCCGTGCA CCTATGTTCA GCTGGATTCA CCGCAGTGCC
      SerAsnSer SerValSer IleIleArgAla ProMetPhe SerTrpIle HisArgSerAla

1381 GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC
      GluPheAsn AsnIleIle ProSerSerGln IleThrGln IleProLeu ThrLysSerThr

1441 AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG
      AsnLeuGly SerGlyThr SerValValLys GlyProGly PheThrGly GlyAspIleLeu

1501 CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC
      ArgArgThr SerProGly GlnIleSerThr LeuArgVal AsnIleThr AlaProLeuSer

1561 CAGCGCTACC GCGTCCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC
      GlnArgTyr ArgValArg IleArgTyrAla SerThrThr AsnLeuGln PheHisThrSer

1621 ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC
      IleAspGly ArgProIle AsnGlnGlyAsn PheSerAla ThrMetSer SerGlySerAsn

1681 CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC
      LeuGlnSer GlySerPhe ArgThrValGly PheThrThr ProPheAsn PheSerAsnGly

1741 AGCAGCGTGT TCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC
      SerSerVal PheThrLeu SerAlaHisVal PheAsnSer GlyAsnGlu ValTyrIleAsp

1801 CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCG AGTACGACCT GGAGAGGGCT
      ArgIleGlu PheValPro AlaGluValThr PheGluAla GluTyrAsp LeuGluArgAla

1861 CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG
      GlnLysAla ValAsnGlu LeuPheThrSer SerAsnGln IleGlyLeu LysThrAspVal

1921 ACCGACTACC ACATCGATCA GGTGAGCAAC CTGGTGGAGT GCTTAAGCGA CGAGTTCTGC
      ThrAspTyr HisIleAsp GlnValSerAsn LeuValGlu CysLeuSer AspGluPheCys

1981 CTGGACGAGA GAAGGAGCT GAGCGAGAAG GTGAAGCACG CCAAGCGCCT GAGCGACGAG
      LeuAspGlu LysLysGlu LeuSerGluLys ValLysHis AlaLysArg LeuSerAspGlu

47/94

FIG. 13 (CONT)

```
2041  CGCAACCTGC TGCAGGACCC CAACTTCCGC GGCATCAACC GCCAGCTGGA CCGCGGCTGG
      ArgAsnLeu LeuGlnAsp ProAsnPheArg GlyIleAsn ArgGlnLeu AspArgGlyTrp

2101  CGAGGCAGCA CCGATATCAC CATCCAGGGC GGCGACGACG TGTTCAAGGA GAACTACGTG
      ArgGlySer ThrAspIle ThrIleGlnGly GlyAspAsp ValPheLys GluAsnTyrVal

2161  ACCCTGCTGG GCACCTTCGA CGAGTGCTAC CCCACCTACC TGTACCAGAA GATCGACGAG
      ThrLeuLeu GlyThrPhe AspGluCysTyr ProThrTyr LeuTyrGln LysIleAspGlu

2221  AGCAAGCTGA AGGCCTACAC CCGCTACCAG CTGCGCGGCT ACATCGAGGA CAGCCAGGAC
      SerLysLeu LysAlaTyr ThrArgTyrGln LeuArgGly TyrIleGlu AspSerGlnAsp

2281  CTGGAAATCT ACCTGATCCG CTACAACGCC AAGCACGAGA CCGTGAACGT GCCCGGCACC
      LeuGluIle TyrLeuIle ArgTyrAsnAla LysHisGlu ThrValAsn ValProGlyThr

2341  GGCAGCCTGT GGCCCCTGAG CGCCCCCAGC CCCATCGGCA AGTGCGGGGA GCCGAATCGA
      GlySerLeu TrpProLeu SerAlaProSer ProIleGly LysCysGly GluProAsnArg

2401  TGCGCTCCGC ACCTGGAGTG GAACCCGGAC CTAGACTGCA GCTGCAGGGA CGGGGAGAAG
      CysAlaPro HisLeuGlu TrpAsnProAsp LeuAspCys SerCysArg AspGlyGluLys

2461  TGCGCCCACC ACAGCCACCA CTTCAGCCTG GACATCGACG TGGGCTGCAC CGACCTGAAC
      CysAlaHis HisSerHis HisPheSerLeu AspIleAsp ValGlyCys ThrAspLeuAsn

2521  GAGGACCTGG GCGTGTGGGT GATCTTCAAG ATCAAGACCC AGGACGGCCA CGCCCGCCTG
      GluAspLeu GlyValTrp ValIlePheLys IleLysThr GlnAspGly HisAlaArgLeu

2581  GGCAATCTAG AGTTCCTGGA GGAGAAGCCC CTGGTGGGCG AGGCCCTGGC CCGCGTGAAG
      GlyAsnLeu GluPheLeu GluGluLysPro LeuValGly GluAlaLeu AlaArgValLys

2641  CGCGCCGAGA AGAAGTGGCG CGACAAGCGC GAGAAGCTGG AGTGGGAGAC CAACATCGTG
      ArgAlaGlu LysLysTrp ArgAspLysArg GluLysLeu GluTrpGlu ThrAsnIleVal

2701  TACAAGGAGG CCAAGGAGAG CGTGGACGCC CTGTTCGTGA ACAGCCAGTA CGACCGCCTG
      TyrLysGlu AlaLysGlu SerValAspAla LeuPheVal AsnSerGln TyrAspArgLeu

2761  CAGGCCGACA CCAACATCGC CATGATCCAC GCCGCCGACA AGCGCGTGCA CAGCATTCGC
      GlnAlaAsp ThrAsnIle AlaMetIleHis AlaAlaAsp LysArgVal HisSerIleArg

2821  GAGGCCTACC TGCCCGAGCT GAGCGTGATC CCCGGCGTGA ACGCCGCCAT CTTCGAGGAA
      GluAlaTyr LeuProGlu LeuSerValIle ProGlyVal AsnAlaAla IlePheGluGlu

2881  CTCGAGGGCC GCATCTTCAC CGCCTTCAGC CTGTACGACG CCCGCAACGT GATCAAGAAC
      LeuGluGly ArgIlePhe ThrAlaPheSer LeuTyrAsp AlaArgAsn ValIleLysAsn

2941  GGCGACTTCA CAACGGCCT GAGCTGCTGG AACGTGAAGG CCACGTGGA CGTGGAGGAG
      GlyAspPhe AsnAsnGly LeuSerCysTrp AsnValLys GlyHisVal AspValGluGlu

3001  CAGAACAACC ACCGCAGCGT GCTGGTGGTG CCCGAGTGGG AGGCCGAGGT GAGCCAGGAG
      GlnAsnAsn HisArgSer ValLeuValVal ProGluTrp GluAlaGlu ValSerGlnGlu
```

48/94

FIG. 13 (CONT)

```
3061  GTGCGCGTGT GCCCCGGCCG CGGCTACATC CTGCGCGTGA CCGCCTACAA GGAGGGCTAC
      ValArgVal  CysProGly  ArgGlyTyrIle LeuArgVal  ThrAlaTyr  LysGluGlyTyr

3121  GGCGAGGGCT GCGTGACCAT CCACGAGATC GAGAACAACA CCGACGAGCT CAAGTTCAGC
      GlyGluGly  CysValThr  IleHisGluIle GluAsnAsn  ThrAspGlu  LeuLysPheSer

3181  AACTGCGTGG AGGAGGAGGT GTACCCCAAC AACACCGTGA CCTGCAACGA CTACACCGCG
      AsnCysVal  GluGluGlu  ValTyrProAsn AsnThrVal  ThrCysAsn  AspTyrThrAla

3241  ACCCAGGAGG AGTACGAGGG CACCTACACC AGCCGCAACC GCGGCTACGA CGGCGCCTAC
      ThrGlnGlu  GluTyrGlu  GlyThrTyrThr SerArgAsn  ArgGlyTyr  AspGlyAlaTyr

3301  GAGAGCAACA GCAGCGTGCC CGCCGACTAC GCCAGCGCCT ACGAGGAGAA GGCCTACACC
      GluSerAsn  SerSerVal  ProAlaAspTyr AlaSerAla  TyrGluGlu  LysAlaTyrThr

3361  GACGGCCGCC GCGACAACCC CTGCGAGAGC AACCGCGGCT ACGGCGACTA CACCCCCCTG
      AspGlyArg  ArgAspAsn  ProCysGluSer AsnArgGly  TyrGlyAsp  TyrThrProLeu

3421  CCCGCCGGCT ACGTGACCAA GGAGCTGGAG TACTTCCCCG AGACCGACAA GGTGTGGATC
      ProAlaGly  TyrValThr  LysGluLeuGlu TyrPhePro  GluThrAsp  LysValTrpIle

3481  GAGATCGGCG AGACCGAGGG CACCTTCATC GTGGACAGCG TGGAGCTGCT GCTGATGGAG
      GluIleGly  GluThrGlu  GlyThrPheIle ValAspSer  ValGluLeu  LeuLeuMetGlu

3541  GAGTAG
      Glu---
```

FIG. 14

FIG. 15

```
  1  ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGAG
     MetAspAsn AsnProAsn IleAsnGluCys IleProTyr AsnCysLeu SerAsnProGlu

 61  GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCCTG
     ValGluVal LeuGlyGly GluArgIleGlu ThrGlyTyr ThrProIle AspIleSerLeu

121  AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCCTG
     SerLeuThr GlnPheLeu LeuSerGluPhe ValProGly AlaGlyPhe ValLeuGlyLeu

181  GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGATC
     ValAspIle IleTrpGly IlePheGlyPro SerGlnTrp AspAlaPhe LeuValGlnIle

241  GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCCTG
     GluGlnLeu IleAsnGln ArgIleGluGlu PheAlaArg AsnGlnAla IleSerArgLeu

301  GAGGGCCTGA GCAACCTGTA CCAAATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGAC
     GluGlyLeu SerAsnLeu TyrGlnIleTyr AlaGluSer PheArgGlu TrpGluAlaAsp

361  CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGCC
     ProThrAsn ProAlaLeu ArgGluGluMet ArgIleGln PheAsnAsp MetAsnSerAla

421  CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGTG
     LeuThrThr AlaIlePro LeuPheAlaVal GlnAsnTyr GlnValPro LeuLeuSerVal

481  TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTCAGCGT GTTCGGCCAG
     TyrValGln AlaAlaAsn LeuHisLeuSer ValLeuArg AspValSer ValPheGlyGln

541  CGCTGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGATC
     ArgTrpGly PheAspAla AlaThrIleAsn SerArgTyr AsnAspLeu ThrArgLeuIle

601  GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG GCCTGGAGCG CGTGTGGGGT
     GlyAsnTyr ThrAspHis AlaValArgTrp TyrAsnThr GlyLeuGlu ArgValTrpGly

661  CCCGACAGCC GCGACTGGAT CAGGTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGTG
     ProAspSer ArgAspTrp IleArgTyrAsn GlnPheArg ArgGluLeu ThrLeuThrVal

721  CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGTG
     LeuAspIle ValSerLeu PheProAsnTyr AspSerArg ThrTyrPro IleArgThrVal

781  AGCCAGCTGA CCCGCGAGAT TTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC
     SerGlnLeu ThrArgGlu IleTyrThrAsn ProValLeu GluAsnPhe AspGlySerPhe

841  CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG
     ArgGlySer AlaGlnGly IleGluGlySer IleArgSer ProHisLeu MetAspIleLeu

901  AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG
     AsnSerIle ThrIleTyr ThrAspAlaHis ArgGlyGlu TyrTyrTrp SerGlyHisGln

961  ATCATGGCCA GCCCCGTCGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC
     IleMetAla SerProVal GlyPheSerGly ProGluPhe ThrPhePro LeuTyrGlyThr
```

FIG. 15 (CONT)

1021 ATGGGCAACG CTGCACCTCA GCAGCGCATC GTGGCACAGC TGGGCCAGGG AGTGTACCGC
      MetGlyAsn AlaAlaPro GlnGlnArgIle ValAlaGln LeuGlyGln GlyValTyrArg

1081 ACCCTGAGCA GCACCCTGTA CCGTCGACCT TTCAACATCG GCATCAACAA CCAGCAGCTG
      ThrLeuSer SerThrLeu TyrArgArgPro PheAsnIle GlyIleAsn AsnGlnGlnLeu

1141 AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG
      SerValLeu AspGlyThr GluPheAlaTyr GlyThrSer SerAsnLeu ProSerAlaVal

1201 TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCTCAGAA CAACAACGTG
      TyrArgLys SerGlyThr ValAspSerLeu AspGluIle ProProGln AsnAsnAsnVal

1261 CCACCTCGAC AGGGCTTCAG CCACCGTCTG AGCCACGTGA GCATGTTCCG CAGTGGCTTC
      ProProArg GlnGlyPhe SerHisArgLeu SerHisVal SerMetPhe ArgSerGlyPhe

1321 AGCAACAGCA GCGTGAGCAT CATCCGTGCA CCTATGTTCA GCTGGATTCA CCGCAGTGCC
      SerAsnSer SerValSer IleIleArgAla ProMetPhe SerTrpIle HisArgSerAla

1381 GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC
      GluPheAsn AsnIleIle ProSerSerGln IleThrGln IleProLeu ThrLysSerThr

1441 AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG
      AsnLeuGly SerGlyThr SerValValLys GlyProGly PheThrGly GlyAspIleLeu

1501 CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC
      ArgArgThr SerProGly GlnIleSerThr LeuArgVal AsnIleThr AlaProLeuSer

1561 CAGCGCTACC GCGTCCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC
      GlnArgTyr ArgValArg IleArgTyrAla SerThrThr AsnLeuGln PheHisThrSer

1621 ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC
      IleAspGly ArgProIle AsnGlnGlyAsn PheSerAla ThrMetSer SerGlySerAsn

1681 CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC
      LeuGlnSer GlySerPhe ArgThrValGly PheThrThr ProPheAsn PheSerAsnGly

1741 AGCAGCGTGT TCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC
      SerSerVal PheThrLeu SerAlaHisVal PheAsnSer GlyAsnGlu ValTyrIleAsp

1801 CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCG AGTACGACCT GGAGAGGGCT
      ArgIleGlu PheValPro AlaGluValThr PheGluAla GluTyrAsp LeuGluArgAla

1861 CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG
      GlnLysAla ValAsnGlu LeuPheThrSer SerAsnGln IleGlyLeu LysThrAspVal

1921 ACCGACTACC ACATCGATCA AGTATCCAAT TTAGTTGAGT GTTTATCTGA TGAATTTTGT
      ThrAspTyr HisIleAsp GlnValSerAsn LeuValGlu CysLeuSer AspGluPheCys

1981 CTGGATGAAA AAAAAGAATT GTCCGAGAAA GTCAAACATG CGAAGCGACT TAGTGATGAG
      LeuAspGlu LysLysGlu LeuSerGluLys ValLysHis AlaLysArg LeuSerAspGlu

FIG. 15 (CONT)

2041   CGGAATTTAC TTCAAGATCC AAACTTTAGA GGGATCAATA GACAACTAGA CCGTGGCTGG
       ArgAsnLeu LeuGlnAsp ProAsnPheArg GlyIleAsn ArgGlnLeu AspArgGlyTrp

2101   AGAGGAAGTA CGGATATTAC CATCCAAGGA GGCGATGACG TATTCAAAGA GAATTACGTT
       ArgGlySer ThrAspIle ThrIleGlnGly GlyAspAsp ValPheLys GluAsnTyrVal

2161   ACGCTATTGG GTACCTTTGA TGAGTGCTAT CCAACGTATT TATATCAAAA AATAGATGAG
       ThrLeuLeu GlyThrPhe AspGluCysTyr ProThrTyr LeuTyrGln LysIleAspGlu

2221   TCGAAATTAA AAGCCTATAC CCGTTACCAA TTAAGAGGGT ATATCGAAGA TAGTCAAGAC
       SerLysLeu LysAlaTyr ThrArgTyrGln LeuArgGly TyrIleGlu AspSerGlnAsp

2281   TTAGAAATCT ATTTAATTCG CTACAATGCC AAACACGAAA CAGTAAATGT GCCAGGTACG
       LeuGluIle TyrLeuIle ArgTyrAsnAla LysHisGlu ThrValAsn ValProGlyThr

2341   GGTTCCTTAT GGCCGCTTTC AGCCCCAAGT CCAATCGGCA AGTGCGGGGA GCCGAATCGA
       GlySerLeu TrpProLeu SerAlaProSer ProIleGly LysCysGly GluProAsnArg

2401   TGCGCTCCGC ACCTGGAGTG GAACCCGGAC CTAGACTGCA GCTGCAGGGA CGGGGAGAAG
       CysAlaPro HisLeuGlu TrpAsnProAsp LeuAspCys SerCysArg AspGlyGluLys

2461   TGCGCCCACC ACAGCCACCA CTTCAGCCTG GACATCGACG TGGGCTGCAC CGACCTGAAC
       CysAlaHis HisSerHis HisPheSerLeu AspIleAsp ValGlyCys ThrAspLeuAsn

2521   GAGGACCTGG GCGTGTGGGT GATCTTCAAG ATCAAGACCC AGGACGGCCA CGCCCGCCTG
       GluAspLeu GlyValTrp ValIlePheLys IleLysThr GlnAspGly HisAlaArgLeu

2581   GGCAATCTAG AATTTCTCGA AGAGAAACCA TTAGTAGGAG AAGCACTAGC TCGTGTGAAA
       GlyAsnLeu GluPheLeu GluGluLysPro LeuValGly GluAlaLeu AlaArgValLys

2641   AGAGCGGAGA AAAAATGGAG AGACAAACGT GAAAAATTGG AATGGGAAAC AAATATTGTT
       ArgAlaGlu LysLysTrp ArgAspLysArg GluLysLeu GluTrpGlu ThrAsnIleVal

2701   TATAAGAGG CAAAAGAATC TGTAGATGCT TTATTTGTAA ACTCTCAATA TGATAGATTA
       TyrLysGlu AlaLysGlu SerValAspAla LeuPheVal AsnSerGln TyrAspArgLeu

2761   CAAGCGGATA CCAACATCGC GATGATTCAT GCGGCAGATA AACGCGTTCA TAGCATTCGA
       GlnAlaAsp ThrAsnIle AlaMetIleHis AlaAlaAsp LysArgVal HisSerIleArg

2821   GAAGCTTATC TGCCTGAGCT GTCTGTGATT CCGGGTGTCA ATGCGGCTAT TTTTGAAGAA
       GluAlaTyr LeuProGlu LeuSerValIle ProGlyVal AsnAlaAla IlePheGluGlu

2881   TTAGAAGGGC GTATTTTCAC TGCATTCTCC CTATATGATG CGAGAAATGT CATTAAAAAT
       LeuGluGly ArgIlePhe ThrAlaPheSer LeuTyrAsp AlaArgAsn ValIleLysAsn

2941   GGTGATTTTA ATAATGGCTT ATCCTGCTGG AACGTGAAAG GGCATGTAGA TGTAGAAGAA
       GlyAspPhe AsnAsnGly LeuSerCysTrp AsnValLys GlyHisVal AspValGluGlu

3001   CAAAACAACC ACCGTTCGGT CCTTGTTGTT CCGGAATGGG AAGCAGAAGT GTCACAAGAA
       GlnAsnAsn HisArgSer ValLeuValVal ProGluTrp GluAlaGlu ValSerGlnGlu

FIG. 15 (CONT)

```
3061  GTTCGTGTCT GTCCGGGTCG TGGCTATATC CTTCGTGTCA CAGCGTACAA GGAGGGATAT
      ValArgVal CysProGly ArgGlyTyrIle LeuArgVal ThrAlaTyr LysGluGlyTyr

3121  GGAGAAGGTT GCGTAACCAT TCATGAGATC GAGAACAATA CAGACGAACT GAAGTTTAGC
      GlyGluGly CysValThr IleHisGluIle GluAsnAsn ThrAspGlu LeuLysPheSer

3181  AACTGTGTAG AAGAGGAAGT ATATCCAAAC AACACGGTAA CGTGTAATGA TTATACTGCG
      AsnCysVal GluGluGlu ValTyrProAsn AsnThrVal ThrCysAsn AspTyrThrAla

3241  ACTCAAGAAG AATATGAGGG TACGTACACT TCTCGTAATC GAGGATATGA CGGAGCCTAT
      ThrGlnGlu GluTyrGlu GlyThrTyrThr SerArgAsn ArgGlyTyr AspGlyAlaTyr

3301  GAAAGCAATT CTTCTGTACC AGCTGATTAT GCATCAGCCT ATGAAGAAAA AGCATATACA
      GluSerAsn SerSerVal ProAlaAspTyr AlaSerAla TyrGluGlu LysAlaTyrThr

3361  GATGGACGAA GAGACAATCC TTGTGAATCT AACAGAGGAT ATGGGGATTA CACACCACTA
      AspGlyArg ArgAspAsn ProCysGluSer AsnArgGly TyrGlyAsp TyrThrProLeu

3421  CCAGCTGGCT ATGTGACAAA AGAATTAGAG TACTTCCCAG AAACCGATAA GGTATGGATT
      ProAlaGly TyrValThr LysGluLeuGlu TyrPhePro GluThrAsp LysValTrpIle

3481  GAGATCGGAG AAACGGAAGG AACATTCATC GTGGACAGCG TGGAATTACT TCTTATGGAG
      GluIleGly GluThrGlu GlyThrPheIle ValAspSer ValGluLeu LeuLeuMetGlu

3541  GAATAAG
      Glu---
```

FIG. 16

FIG. 17

FIG. 18

PCR fragment replacing Pith Nco I (ATG start)
with Bam HI. Pith-5'-GGATCCAACA(ATG)-3'-syn.-Bt
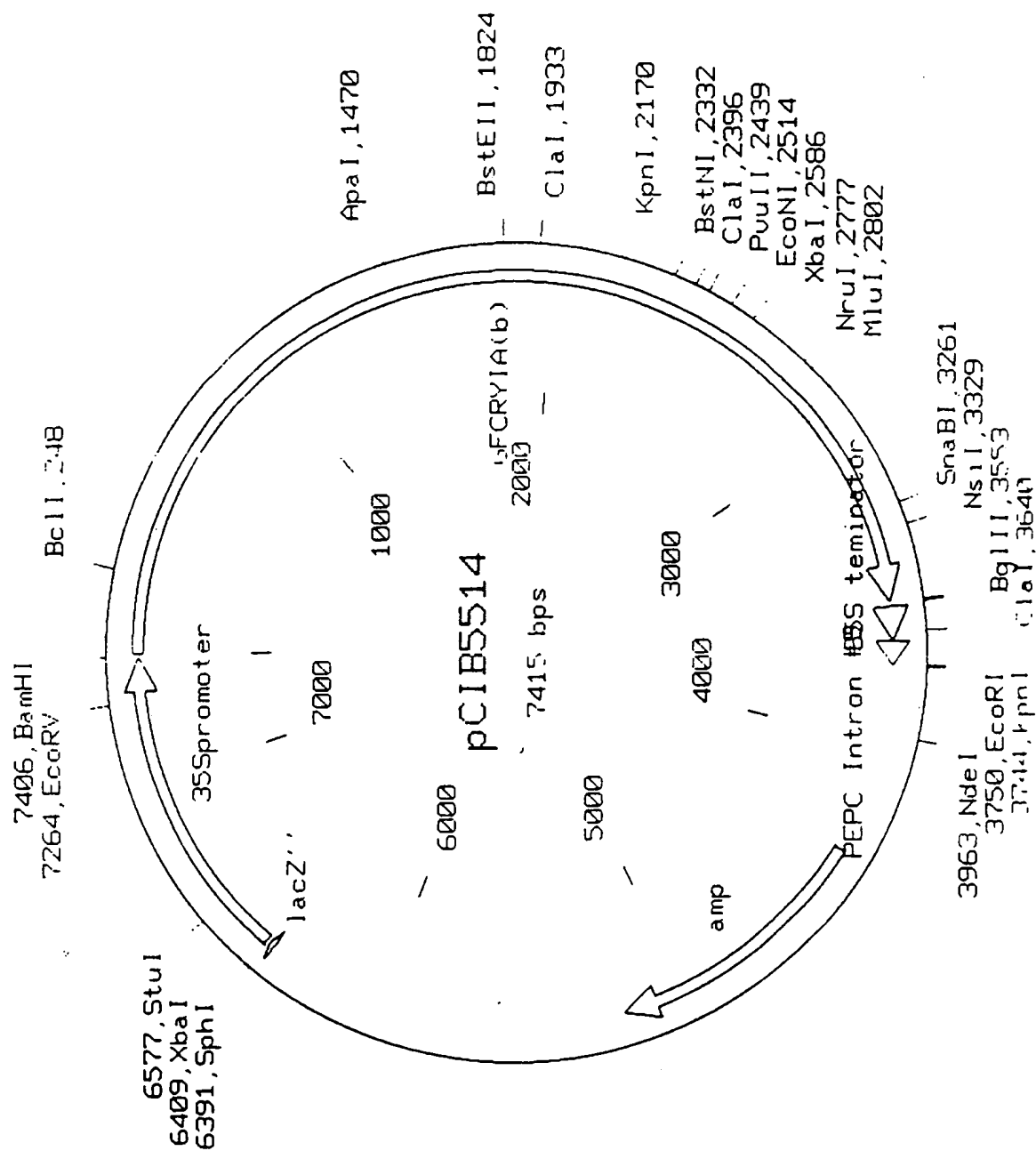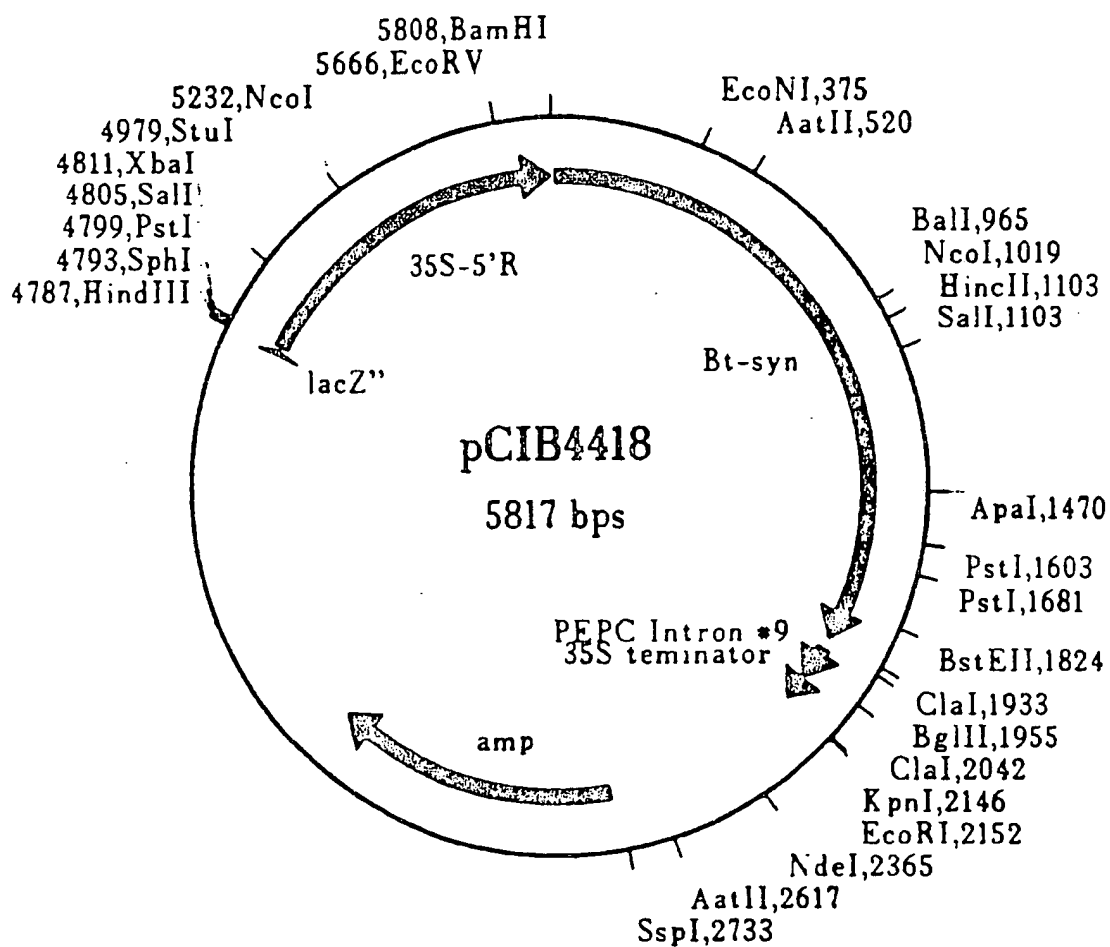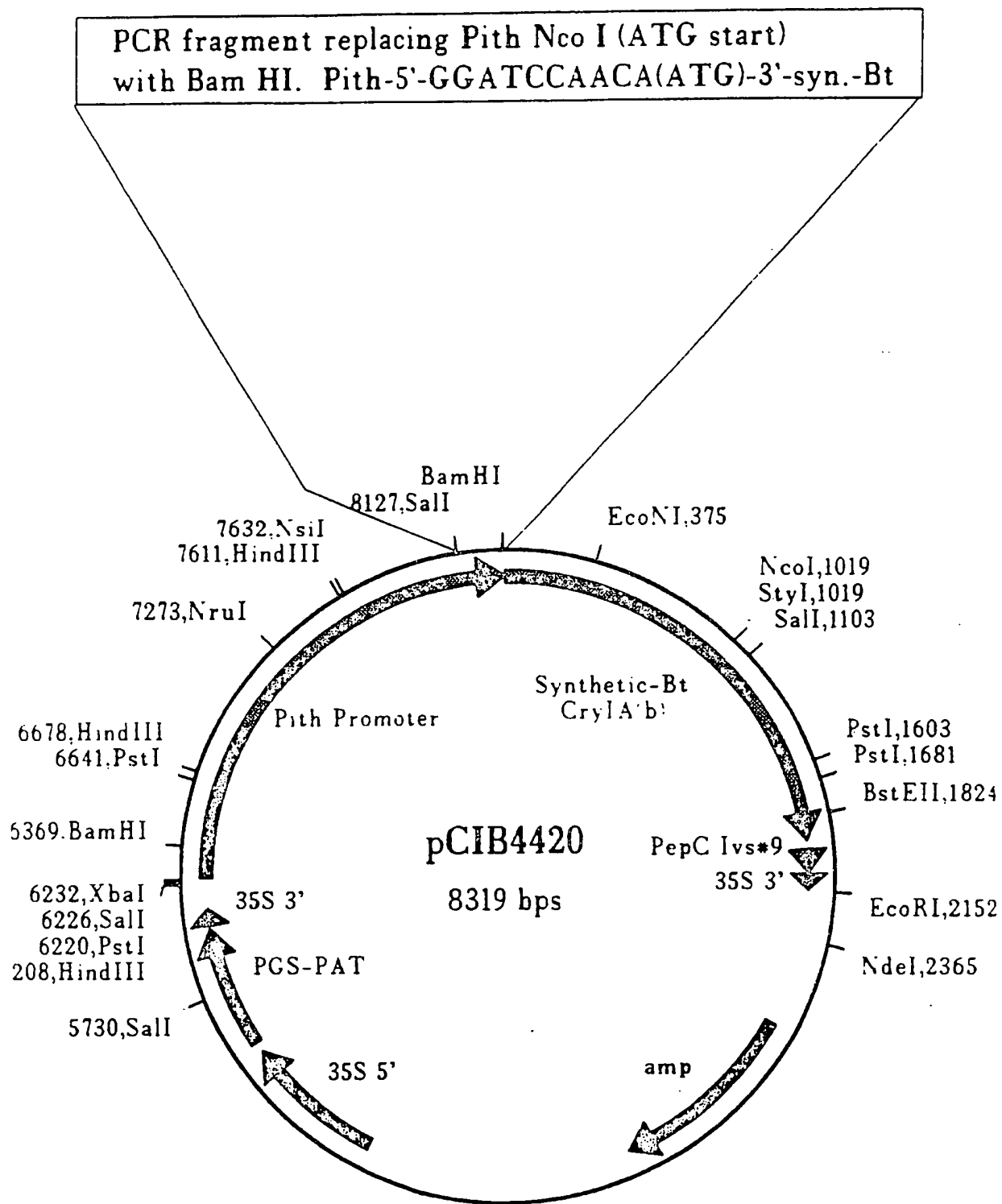
FIG. 19
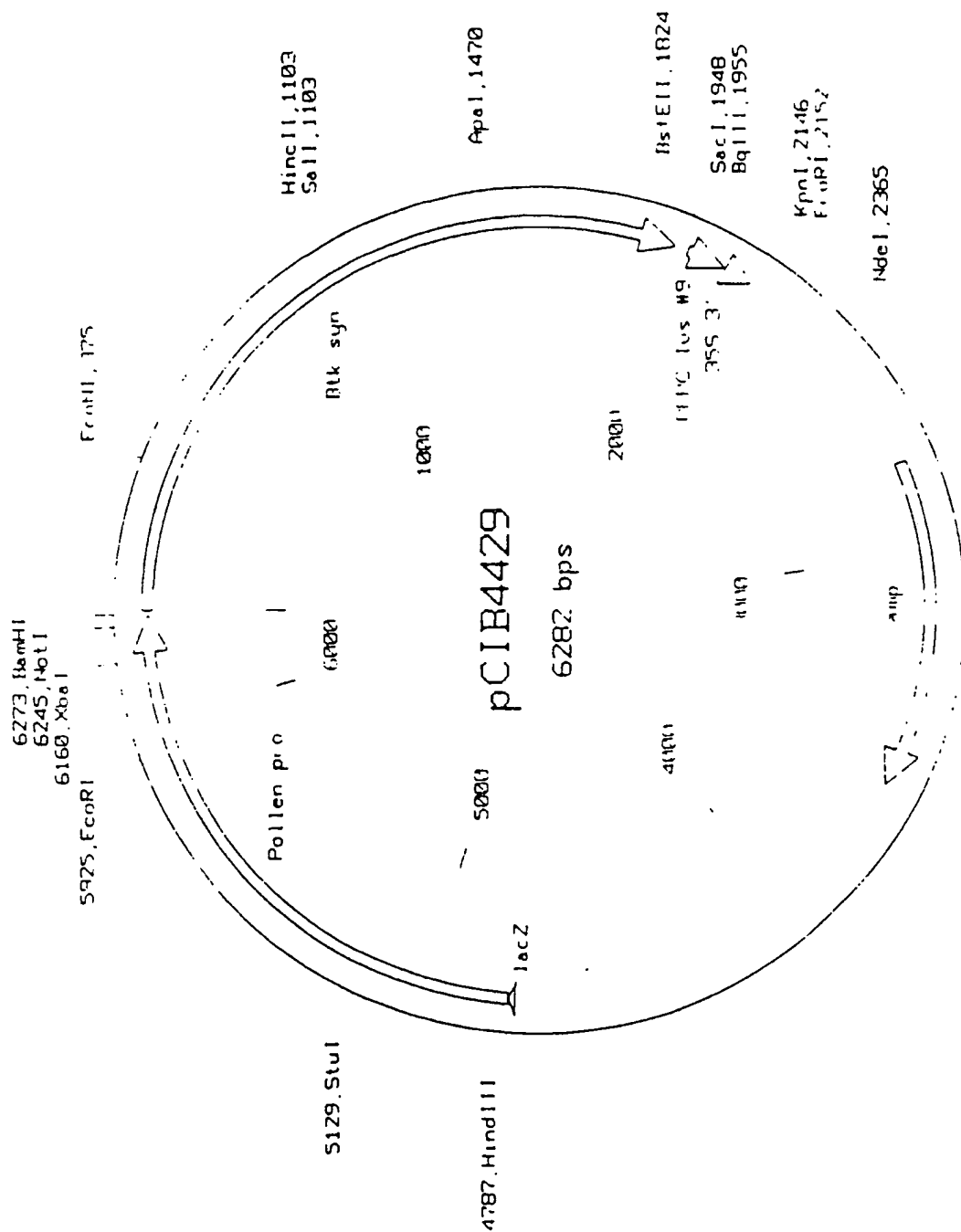
FIG. 20
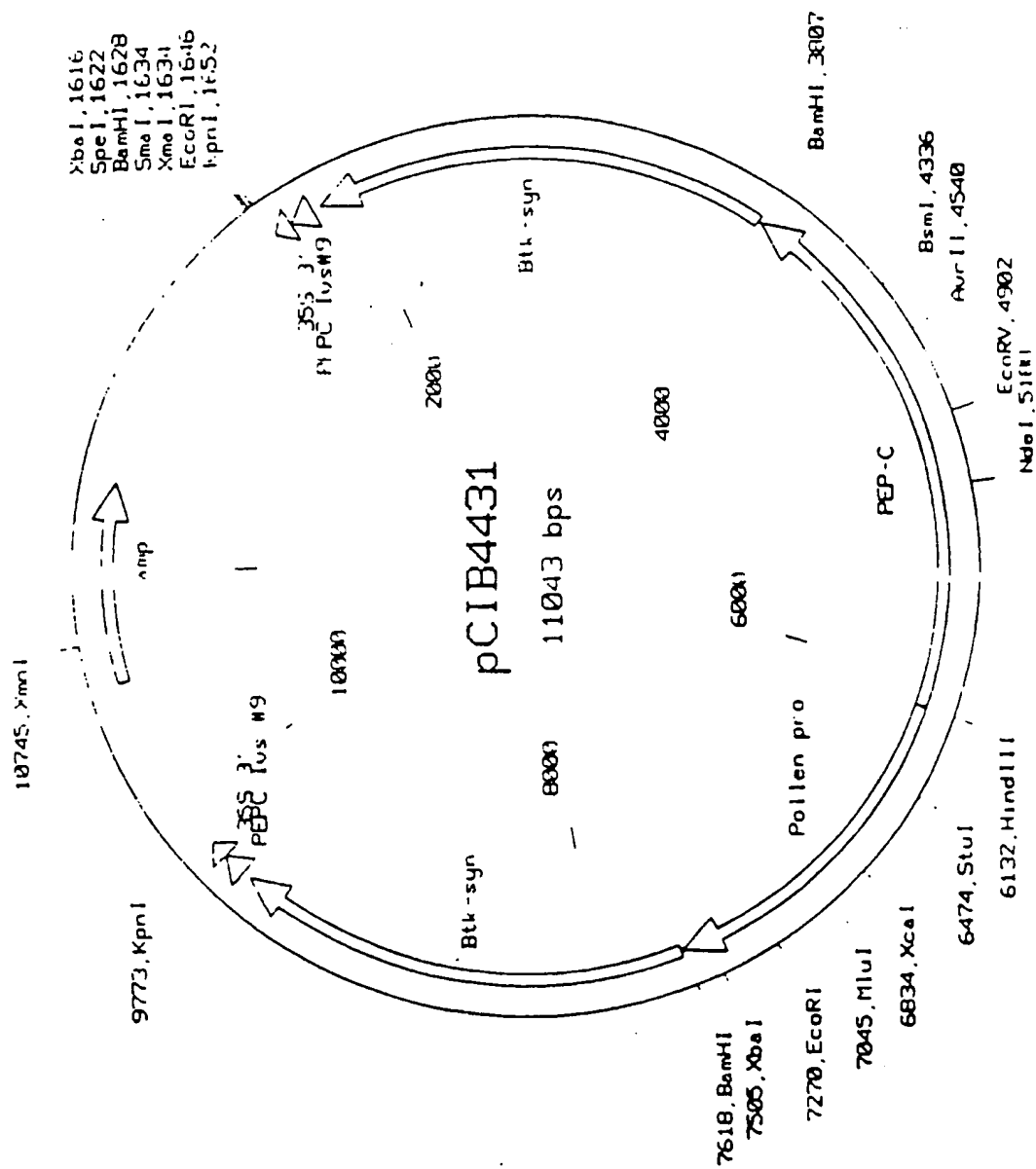
FIG. 21
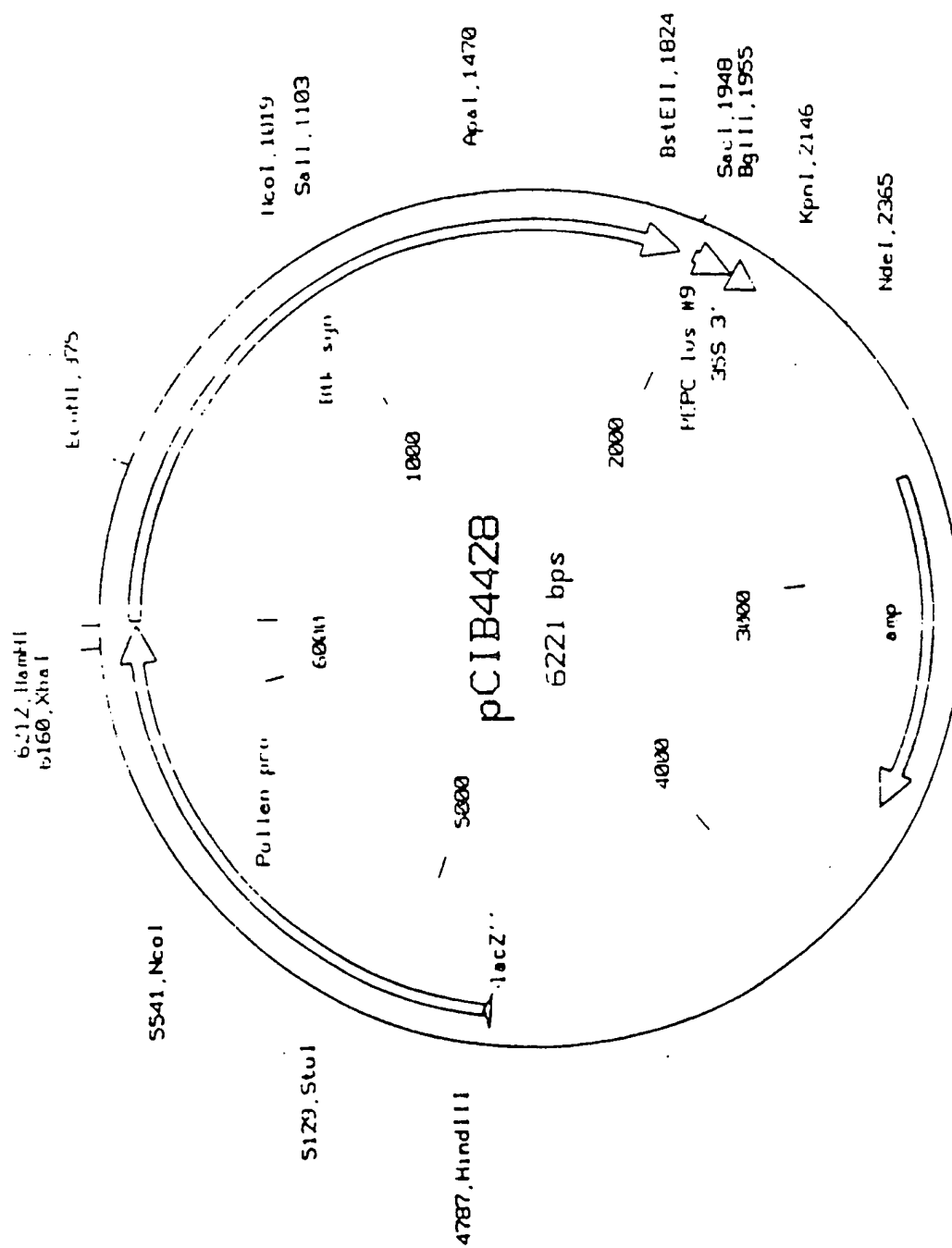
FIG. 22
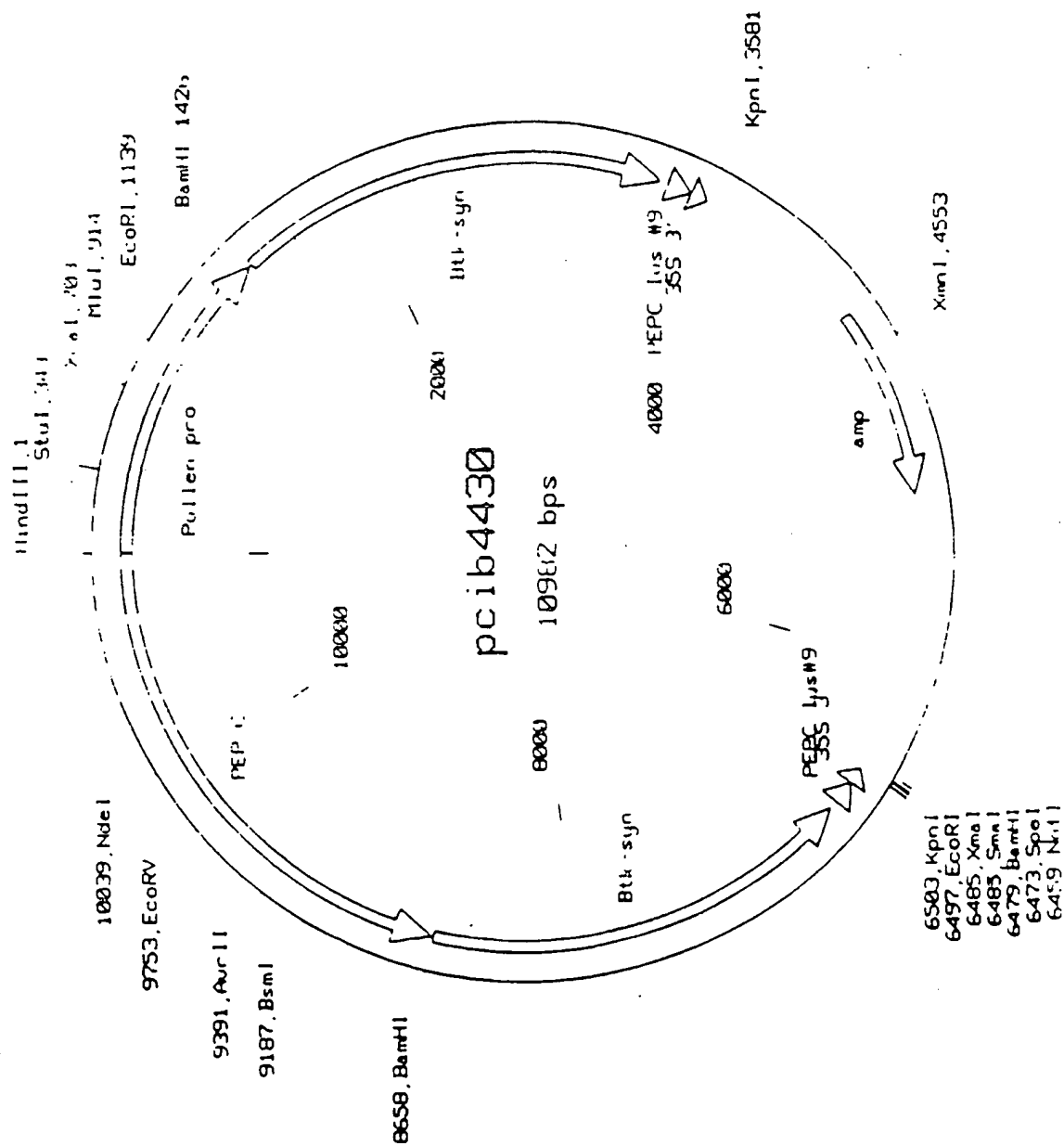
FIG. 23A

CrylA(b) Protein Levels in Transgenic Maize

ELISA Bt Values of Field Plants:

INBRED X PARENT	ABRU PLANT Number	ng Bt/mg protein
2ND01X171-4A	1646	29
5N984X171-4A	857	1705
5N984X171-4A	870	1760
5N984X171-13	969	22
5N984X171-15	1468	17
5N984X171-15	1470	28
5N984X171-14A	1502	180
5N984X171-14A	1529	1500
5N984X176-11	1667	408
5N984X176-11	1671	1270
5N984X176-11	1673	1522
5N984X176-11	1675	943
5N984X176-11	1679	967
5N984X171-4B	1942	15
5N984X171-4B	1946	16
5NA56X171-16ABX	1101	30
5NA89X176-11	1622	959
5NA89X176-11	1630	1172
5NA89X176-11	1635	1100
6F010X171-4	825	103
6F010X171-4	832	1298

·Bt levels are in ng crylA(b) mg total protein

·Data are from progeny of the described maize transformants expressing the crylA(b) protein.

·ELISA analysis of transgenic plant material was carried out using standard procedures as described elsewhere.

FIG. 23B

Bioassay of European corn borer, Ostrinia nubilalis, and sugarcane borer, Diatraea saccharalis.

Plasmid	Promoter	Cross	Plant No.	Bt Gene	Percent Mortality	
					Ostrinia	Diatraea
pCIB4431	PEPC	5N984 X 176-8B	21	+	100	100
			22		0	0
			40	+	100	100
pCIB4431	PEPC	5N984 X 176-11	95	+	100	100
			96	-	0	0
			98	+	100	100
pCIB4418	35S	5N984 X 171-14A	45	-	0	10
			64	+	100	90
			68	+	100	100
pCIB4431	PEPC	2N217AF X 176-8B	1	-	0	0
			3	+	100	100
			4	+	100	100
pCIB4418	35S	2N217AF X 171-15	70	-	10	0
			83	+	90	80
			88	+	90	100

FIG. 23C

CryIA(b) Protein Levels in Transgenic Maize

Greenhouse plants

35S LINE	LEAF	PITH	ROOT	POLLEN
6F010 x 171-4A	-409 + 288	NT	NT	NT
5N984 x 171-14A	256 + 159	191	198	30
6F010 x 171-16AB	240 + 174	221	271	NT
5N984 x 171-13	201 + 94	NT	NT	NT
5NA89 x 171-13	37 + 7	150	0	NT
5N984 x 171-18	7.7 + 3	NT	NT	NT
6N615 x 171-16AB	7.5 + 3	0	0	
PEPC LINE				
6N615 x 176-11	1126 + 419	41	19	NT
6F010 x 176-10	774 + 159	NT	NT	130
5N984 x 176-11	719 + 128	16	20	186

·Bt levels are in ng cryIA(b)/mg total protein

Data are from progeny of the described maize transformants expressing the cryIA(b) protein.

ELISA analysis of transgenic plant material was carned out using standard procedures as described elsewhere.

FIG. 23D

Bioassay of European corn borer. Ostrinia nubilalis. on Pith:SynBt maize

Plasmid	Promoter	Event	Plant No.	Percent Mortality
pCIB4433	Pith	JS21A-Top	1	90
			3	80
			11	90
			13	70
			14	75
			19	85
			28	80
pCIB4433	Pith	JS22D-Mid	3	70
			4	65
			7	85
			17	95
		Control	1	5
			2	0
			3	0

FIG. 23E

EXPRESSION OF THE CRYIA(b) GENE IN TRANSGENIC MAIZE USING THE PITH-PREFERRED PROMOTER

Leaf samples from small plantlets transformed with pCIB4433 using procedures described elsewhere were analyzed for the presence of the cryIA(b) protein using ELISA. All plants expressing cryIA(b) were found to be insecticidal in the standard European corn borer bioassay.

Note that the pith-preferred promoter has a low, but detectable level of expression in leaf tissue of maize. Detection of CryIA(b) protein is consistent with this pattern of expression.

PLANT NUMBER	ng cyIA(b)/mg protein
JS21A-1 TOP	169
JS21A-2 TOP	0
JS21A-3 TOP	113
JS21A-11 TOP	127
JS21A-12 TOP	112
JS21A-13 TOP	97
JS21A-14 TOP	118
JS21A-19 TOP	82
JS21A-24 TOP	0
JS21A-28 TOP	154
JS22D-3 MID	2946
JS22D-4 MID	5590
JS22D-11 MID	215
JS22D-17 MID	3004

Figure 1000 23E

FIG. 24

```
   1 GAATTCGGATCCATTAAAGAAGTCTTTGAACAGATTCTAGAGATCTAGTTTAATGAGCTC    60
  61 CCAAAAGTCTTGAAAAAATTCAGCGGGGAGGCCATTAGGGCAGGGGTACTGTTATGTTTT   120
 121 AAAGAGAACACCACTTTCTTGATCTCTTCTAAAGAGAAATGTTTTGTAAGAAGGATCCTG   180
 181 TCCTCCTCATCCAACCTTTTCATCGGCAAATTTTTCATAGAGATATTAGAGGCAAGAGAG   240
 241 GGGCCAAAAAGATCCATGTAAATGGAAGTGGCCACCTGGTTGATACCTCCCTCATCTTCA   300
 301 ACAGAAAATCCATTATGAAAAAGTGAATGGATTTTAAACTCTTCTTTTTCTTCCCTTTTG   360
 361 CAATGAGCTGAAAATATCTGGTATTATTCTCATCACCCTCATTAATGAATCTGTCCCTAG   420
 421 CAATTTGCTTTCTCTTGATCCCTTCTGCAGCCACCATGTTTCTTAAATTCCACTCCATAT   480
 481 CAAGCTTTTCCAATCTATCAGAATCTGAGATGGCTGCAATCTCTCTCATTTTCTCAAGGA   540
 541 TATCGATGTTATCCATAAGGTATTTCTTGAACTTCTTATATTTCCCTTCGACATTTATAT   600
 601 TCCATCCTTTCAACATTTTTTTGTTCAATCTTTTTTGTTTTTTTCCTTTCCAAACATCGA   660
 661 TACATTTCCTGCTCCTCACAGGTAAGGACGAGCTTTCAAAAAACCTTCTGCTTTAAAGTC   720
 721 AGGTCTGAGCCTCCAGCAAAGCTCACATATCTAAAGTCCCTCTTCTTAGTTGGGACAGAG   780
 781 TCAGTGCTAAGACACATGGGAACATGACCAGAAAAAAAAAATCATATTTAGCCCAGAGAC   840
 841 AACAATATTCTTGTACTGCAAGTCTCGTTATGGGCTAGCAAAGGAATCTACCCAACTTCT   900
 901 CAAATGTGTTGGGATGTCAAGTATATAGACTATTCATCAGTTCCAACTCTATCAAACTGT   960
 961 GCAGCTCAATTATAGAGTTGAATAAAGTGCTCCATCTATTTGTTCTTATCCTCATATTTG  1020
1021 GTTAAGATATTAAAATCACCTCCCACCAACATTTAAAGTGCACCATTTAAAGTGGCTCGC  108C
1081 GAGCACCAAACCGCTGAAAACCGGAAATGTTTAGCACGTTGGCAGCGGGACCCTTTTCTA  1140
1141 TCTCATCGTGTTCTTCGTTGTCCACCACGGCCCACGGGCCAACGCTCCTCCATCCTGTAG  1200
1201 TGTAGAGTATATTCCATTTGCGACCGAGCCGAGCATCGATCCAGCCACACTGGCCACTGC  1260
                                                84
1261 CAGCCAGCCATGTGGCACTCCTACGTATACTACGTGAGGTGAGATTCACTCACATGGGAT  1320
-465                                                             -405

1321 GGGACCGAGATATTTTACTGCTGTGGTTGTGTGAGAGATAATAAAGCATTTATGACGATT  138C
1381 GCTGAACAGCACACACCATGCGTCCAGATAGAGAAAGCTTTCTCTCTTTATTCGCATGCA  1440

1441 TGTTTCATTATCTTTTATCATATATATATAACACATATTAAATGATTCTTCGTTCCAATT  1500
-285                                                             -226

1501 TATAATTCATTTGACTTTTTTATCCACCGATGCTCGTTTTATTAAAAAAAAATATTATAAT  1560
-225                                                             -166

1561 TATTGTTACTTTTTGTTGTAATATTGTTTAGCATATAATAAACTTTGATACTAGTATGTT  1620
-165                                                             -106
                                                49
1621 TCCGAGCAAAAAAAAAATATTAATATTTAGATTACGAGCCCATTAATTAATTATATTCGAG  1680
-105                                                             -46

                    83                              +1
1681 ACAAGCGAAGCAAAGCAAAGCAAGCTAATGTTGCCCCTGCTGTGCATGCAGAGGCCCGCT  1740
-45                                                              +15

                                                    73*******
1741 CTTGCTATAAACGAGGCAGCTAGACGCGACTCGACTCATCAGCCTCATCAACCTCGACGA  1800
+16                                                              +75

     ************                                 ↓
1801 AGGAGGAACGAACGGACAGGTTGTTGCACAGAAGCGACATGGCTTTCGCGCCCAAAACGT  1860
+76                                     M  A  F  A  P  K  T  S   +135
```

38

FIG. 24

```
1861  CCTCCTCCTCCTCGCTGTCCTCGGCGTTGCAGGCAGCTCAGTCGCCGCCGCTGCTCCTGA   192?
+126    S   S   S   S   L   S   S   A   L   Q   A   A   Q   S   P   P   L   L   L   R   +195

                                                        40 + 41
1921  GGCGGATGTCGTCGACCGCAACACCGAGACGGAGGTACGACGCGGCCGTCGTCGTCACTA   198?
+196    R   M   S   S   T   A   T   P   R   R   R   Y   D   A   A   V   V   V   T   T   +255

1981  CCACCACCACTGCTAGAGCTGCGGCGGCTGCTGTCACGGTTCCCGCCGCCCCGCCGCAGG   204?
+256    T   T   T   A   R   A   A   A   A   A   V   T   V   P   A   A   P   P   Q   A   +315

                          75                                    $
2041  CGGGCCGCCGCCGCCGGTGCCACCAAAGCAAGCGGCGGCACCCGCAGAGGAGGAGCCGTC   2100
+316    G   R   R   R   R   C   H   Q   S   K   R   R   H   P   Q   R   R   S   R   P   +375

2101  CGGTGTCGGACACCATGGCGGCGCTCATGGCCAAGGGCAAGGTTCGTATAGTACGCGCGC   2160
+376    V   S   D   T   M   A   A   L   M   A   K   G   K

2161  GTGTCGTCGTCGTTATTTTGCGCATAGGCGCGGACATACACGTGCTTTAGCTAGCTAACA   222?
2221  GCTAGATCATCGGTGCAGACGGCGTTCATCCCGTACATCACCGCCGGCGACCCGGACCTA   228?
                    T   A   F   I   P   Y   I   T   A   G   D   P   D   L

2281  GCGACGACGGCCGAGGCGCTGCGTCTGCTGGACGGCTGTGGCGCCGACGTCATCGAGCTG   234?
        A   T   T   A   E   A   L   R   L   L   D   G   C   G   A   D   V   I   E   L

2341  GGGGTACCCTGCTCGGACCCCTACATCGACGGGCCCATCATCCAGGCGTCGGTGGCGCGG   2400
        G   V   P   C   S   D   P   Y   I   D   G   P   I   I   Q   A   S   V   A   R

2401  GCTCTGGCCAGCGGCACCACCATGGACGCCGTGCTGGAGATGCTGAGGGAGGTGACGCCG   246?
        A   L   A   S   G   T   T   M   D   A   V   L   E   M   L   R   E   V   T   P

2461  GAGCTGTCGTGCCCCGTGGTGCTCCTCTCCTACTACAAGCCCATCATGTCTCGCAGCTTG   252?
        E   L   S   C   P   V   V   L   L   S   Y   Y   K   P   I   M   S   R   S   L

2521  GCCGAGATGAAAGAGGCGGGGGGTCCACGGTAACTATAGCTAGCTCTTCCGATCCCCCTTC   258?
        A   E   M   K   E   A   G   V   H

2581  AATTAATTAATTTATAGTAGTCCATTCATGTGATGATTTTTGTTTTTCTTTTTACTGACA   2640
2641  GGTCTTATAGTGCCTGATCTCCCGTACGTGGCCGCGCACTCGCTGTGGAGTGAAGCCAAG   2700
        G   L   I   V   P   D   L   P   Y   V   A   A   H   S   L   W   S   E   A   K

2701  AACAACAACCTGGAGCTGGTAGGTTGAATTAAGTTGATGCATGTGATGATTTATGTAGCT   2760
        N   N   N   L   E   L

2761  AGATCGAGCTAGCTATAATTAGGAGCATATCAGGTGCTGCTGACAACACCAGCCATACCA   2820
                                        V   L   L   T   T   P   A   I   P

2821  GAAGACAGGATGAAGGAGATCACCAAGGCTTCAGAAGGCTTCGTCTACCTGGTAGTTATA   2880
        E   D   R   M   K   E   I   T   K   A   S   E   G   F   V   Y   L

2881  TGTATATATAGATGGACGACGTAACTCATTCCAGCCCCATGCATATATGGAGGCTTCAAT   2940
2941  TCTGCAGAGACGACGAAGACCACGACGACGACTAACACTAGCTAGGGGCGTACGTTGCAG   3000

3001  GTGAGCGTGAACGGAGTGACAGGTCCTCGCGCAAACGTGAACCCACGAGTGGAGTCACTC   3060
        V   S   V   N   G   V   T   G   P   R   A   N   V   N   P   R   V   E   S   L
```

FIG. 24

```
3061  ATCCAGGAGGTTAAGAAGGTGACTAACAAGCCCGTTGCTGTTGGCTTCGGCATATCCAAG  3120
       I  Q  E  V  K  K  V  T  N  K  P  V  A  V  G  F  G  I  S  K

3121  CCCGAGCACGTGAAGCAGGTACGTACGTAGCTGACCAAAAAAAACTGTTAACAAGTTTTG  3180
       P  E  H  V  K

3181  TTTGACAAGCCGGCTACTAGCTAGCTAACAGTGATCAGTGACACACACACACACACAGAT  3240
                                                               Q  I

3241  TGCGCAGTGGGGCGCTGACGGGGTGATCATCGGCAGCGCCATGGTGAGGCAGCTGGGCGA  3300
       A  Q  W  G  A  D  G  V  I  I  G  S  A  M  V  R  Q  L  G  E

3301  AGCGGCTTCTCCCAAGCAAGGCCTGAGGAGGCTGGAGGAGTATGCCAGGGGCATGAAGAA  3360
       A  A  S  P  K  Q  G  L  R  R  L  E  E  Y  A  R  G  M  K  N
                         +++

3361  CGCGCTGCCATGAGTCCATGACAAAGTAAAACGTACAGAGACACTTGATAATATCTATCT  3420
       A  L  P

3421  ATCATCTCGGAGAAGACGACCGACCAATAAAAATAAGCCAAGTGGAAGTGAAGCTTAGCT  3480

3481  GTATATACACCGTACGTCGTCGTCGTCGTTCCGGATCGATCTCGGCCGGCTAGCTAGCAG  3540
3541  AACGTGTACGTAGTAGTATGTAATGCATGGAGTGTGGAGCTACTAGCTAGCTGGCCGTTC  360?
3601  ATTCGATTATAATTCTTCGCTCTGCTGTGGTAGCAGATGTACCTAGTCGATCTTGTACGA  366C

3661  CGAAGAAGCTGGCTAGCTAGCCGTCTCGATCGTATATGTACTGATTAATCTGCAGATTGA  3720
                                                                    $
3721  ATAAAAACTACAGTACGCATATGATGCGTACGTACGTGTGTATAGTTTGTGCTCATATAT  3780
3781  GCTCCTCATCACCTGCCTGATCTGCCCATCGATCTCTCTCGTACTCCTTCCTGTTAAATG  3840
3841  CCTTCTTTGACAGACACACCACCACCAGCAGCAGTGACGCTCTGCACGCCGCCGCTTTAA  3900
3901  GACATGTAAGATATTTTAAGAGGTATAAGATACCAAGGAGCACAAATCTGGAGCACTGGG  396?
3961  ATATTGCAAAGACAAAAAAAAAACAAAATTAAAGTCCCACCAAAGTAGAGATAGTAAAGA  4020
4021  GGTGGATGGATTAAAATTATCTCATGATTTTTGGATCTGCTCAAATAGATCGATATGGTA  4080
4081  TTCAGATCTATGTTGTATAGCCTTTTCATTAGCTTTCTGAAAAAAAAATGGTATGATGAG  4140
4141  TGCGGAGTAGCTAGGGCTGTGAAGGAGTCGGATGGGCTTCCACGTACTTGTTTGTGGCCC  4200
4201  TAGTCCGGTTCTATTTAGGTCCGATCCGAGTCCGGCATGGTCCGGTTCCATACGGGCTAG  4260
4261  GACCAAGCTCGGCACGTGAGTTTTAGGCCCGTCGGCTAGCCCGAGCACGACCCGTTTTTA  4320
4321  AACTGGCTAGGACTCGCCCATTTAATAAGACAAACATTGCAAAAAATAGCTCTATTTTTT  4380
4381  ATTTAAAATATATTGTTTATTTGTGAAATGTGTATTATTTGTAATATATATTATTGTATA  4440
4441  TAGTTATATCTTCAATTATGATTTATAAATATGTTTTTTATTATGAACTCAATTTTAAGT  4500
4501  TTGATTTATGCGTTGGCGGGCTCGAGGAGGCACGGTGAACATTTTTGGGTCGGGCTTAAC  4560
4561  GGGTCGGCCCGGCCCGGTTCGGCCCATCCACGGCCCATCCCGTGTCGGCCTCGTTCGGTG  4620
4621  AGTTCAGCCCGTCGGACAACCCGTCCCCGGCCCGGATAATTAATCGGGCCTAACCGTGGC  4680
4681  GTGCTTAAACGGTCCGTGCCTCAACGGACCGGGCCGCGGGCGGCCCGTTTGACATCTCTA  4740
4741  GTGGTGTGATTAGAGATGGCGATGGGAACCGATCACTGATTCCGTGTGGAGAATTCGATA  4800
4801  TCAAGCTTATCGATACC                                            4817
```

Figure 24.

Entire sequence of the maize TrpA gene, with introns
and exons, transcription and translation strats, start and
stop of cDNA.
$ = start and end of cDNA; +1 = transcription start; 73******* =
primer extension primer; ⌐ = start of translation; +++ = stop
codon; _____ = CCAAT Box, TATAA Box, poly A addition site.
⌐ above underlined sequences are PCR primers.

FIG. 25A

211D   5N984

                60d   70d

P   C BR ES LP MP UP   C BR ES LP MP UP

        211D   5N984

                            39d   46d

S   L   R   P   SH   LP UP   LP MP UP

Northern blot showing differential expression of
TrpA gene in maize tissues. 2 hour exposure against
film at -80C with Dupont Cronex intensifying screens.

FIG. 25B

211D _____ 5N984 _____

|_____ 60d _____|_____ 70d _____|

P C BR ES LP MP UP C BR ES LP MP UP

| 211D | 5N984 |

|_____|_____|

| 39d | 46d |

S L R P SH LP UP LP MP UP
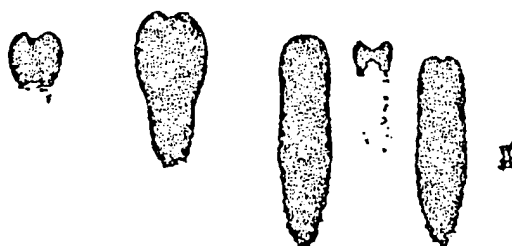
Northern blot showing differential expression of
TrpA gene in maize tissues. 4 hour exposure against
film at -80C with Dupont Cronex intensifying screens.

FIG. 25C

211D	5N984											
	60d						70d					
P	C	BR	ES	LP	MP	UP	C	BR	ES	LP	MP	UP

211D					5N984				
					39d		46d		
S	L	R	P	SH	LP	UP	LP	MP	UP
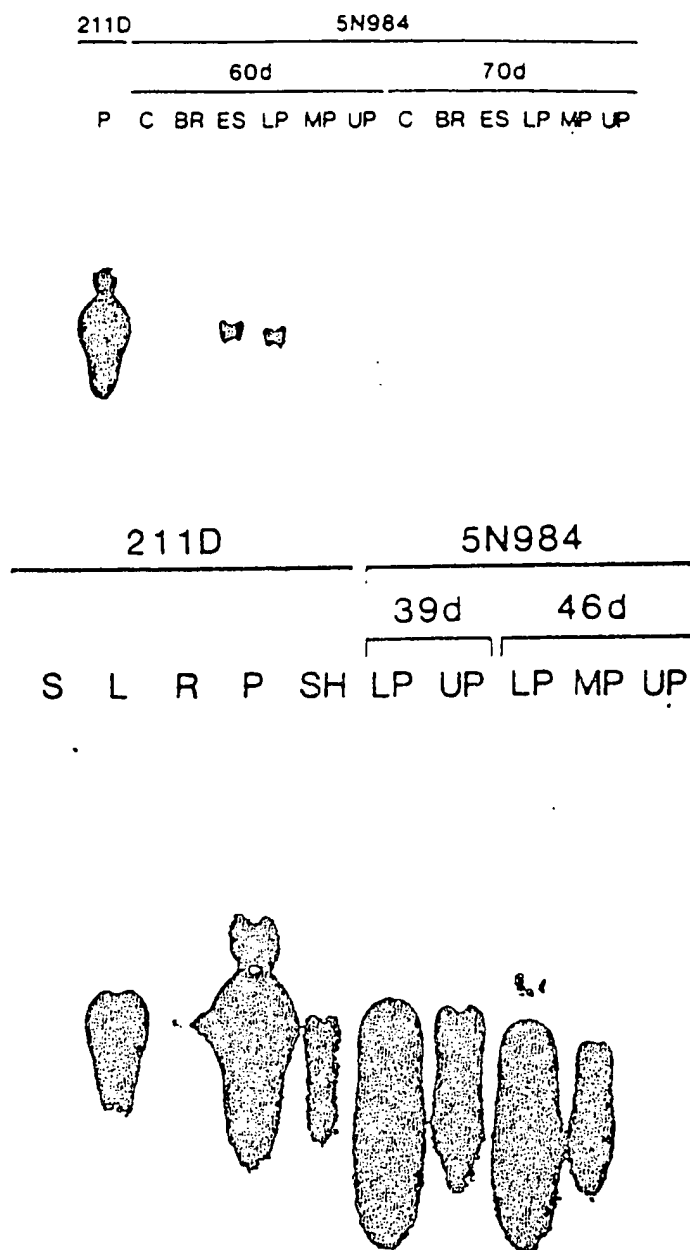
Northern blot showing differential expression of
TrpA gene in maize tissues. 18 hour exposure against
film at -80C with Dupont Cronex intensifying screens.

FIG. 25D

211D	5N984										
	60d						70d				
P	C	BR	ES	LP	MP	UP	C	BR	ES	LP	MP	UP

211D					5N984				
					39d		46d		
S	L	R	P	SH	LP	UP	LP	MP	UP
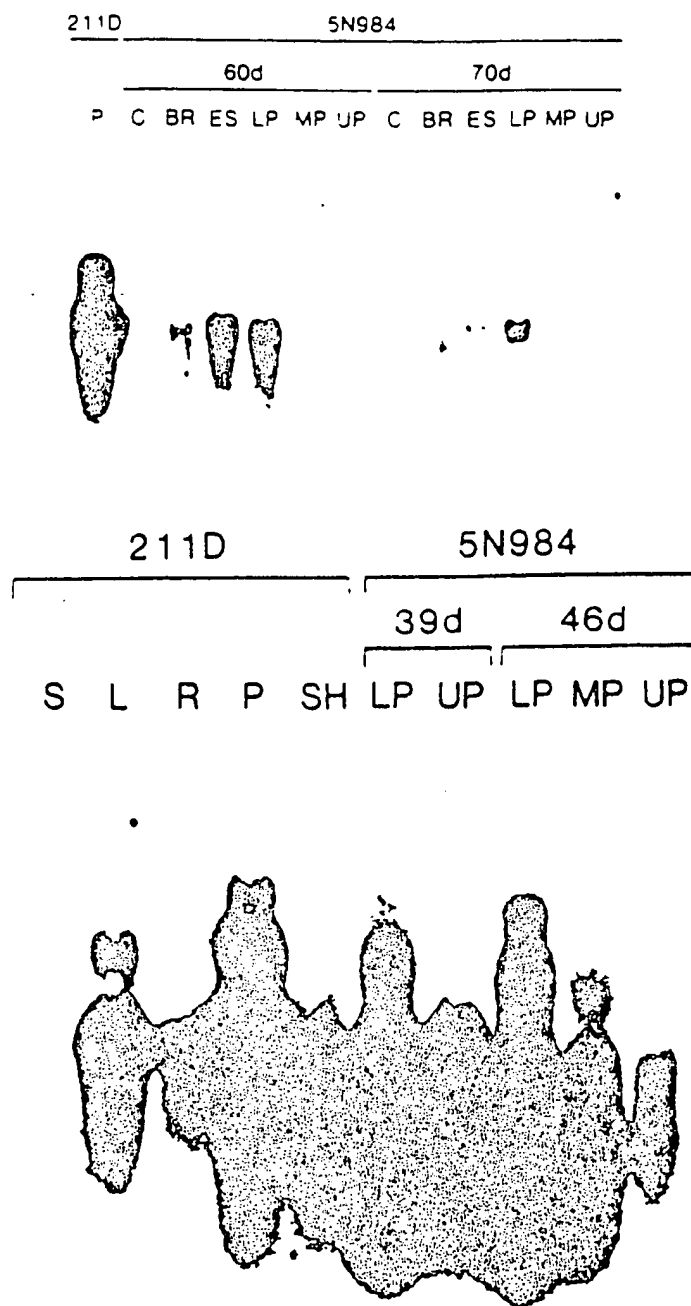
Northern blot showing differential expression of
TrpA gene in maize tissues. 48 hour exposure against
film at -80C with Dupont Cronex intensifying screens.

FIG. 26

```
          211D 5N984          211D
          ―――― ―――――      ――――――――――――――
          L   P   L   P    S1 S2 S3 S4 S5
```

Northern blot showing maize TrpA gene expression in
Funk lines 211D and 5N984 leaf and pith and the absense
of expression in 211D seed   total RNA.
65 hour exposure against film at -80C with Dupont Cronex
intensifying screens.

FIG. 27
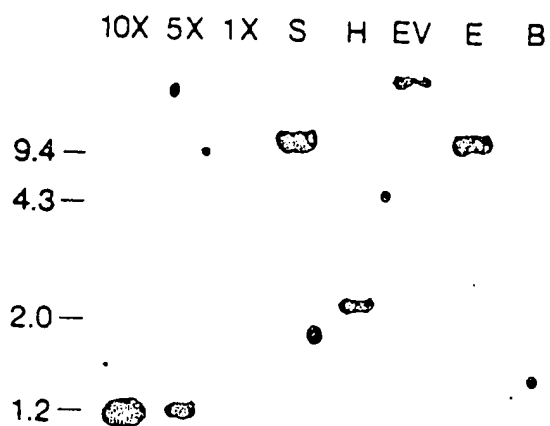
Genomic southern of Funk line 211D probed with the
TrpA cDNA 8-2. B = BamHI, E = EcoRI, EV = EcoRV,
H = HindIII and S = SacI.
120 hour exposure against film at-80C with Dupont Cronex
intensifying screens.

FIG. 28A
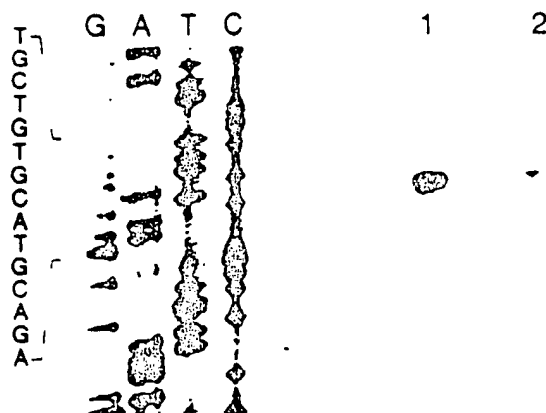
Primer extension showing the transcription start of TrpA
gene and sequencing ladder.
1 hour exposure against film at-80C with Dupont Cronex
intensifying screens.

FIG. 28B
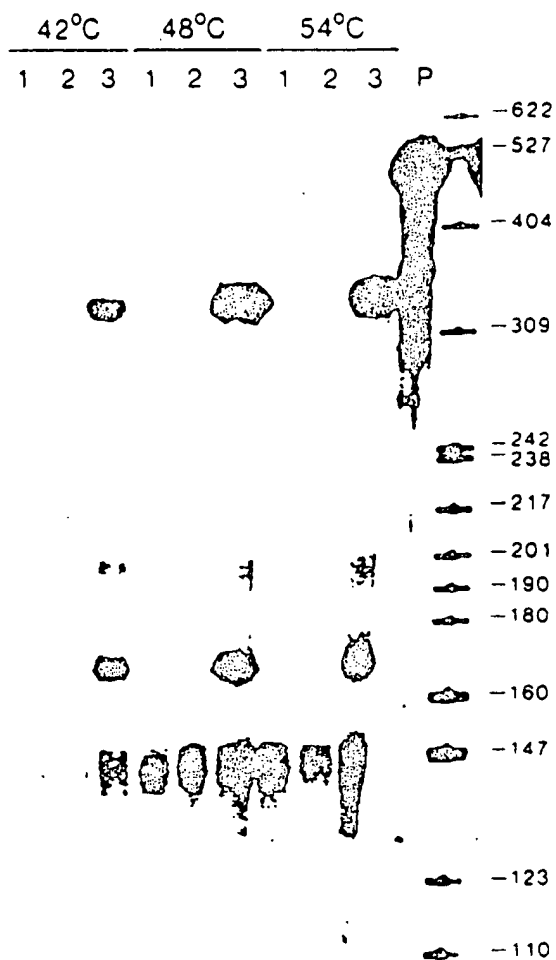
RNase protection of region from +2 bp to +387 bp
with three annealing temperures.
16 hour exposure against film at-80C with Dupont Cronex
intensifying screens.

FIG. 29
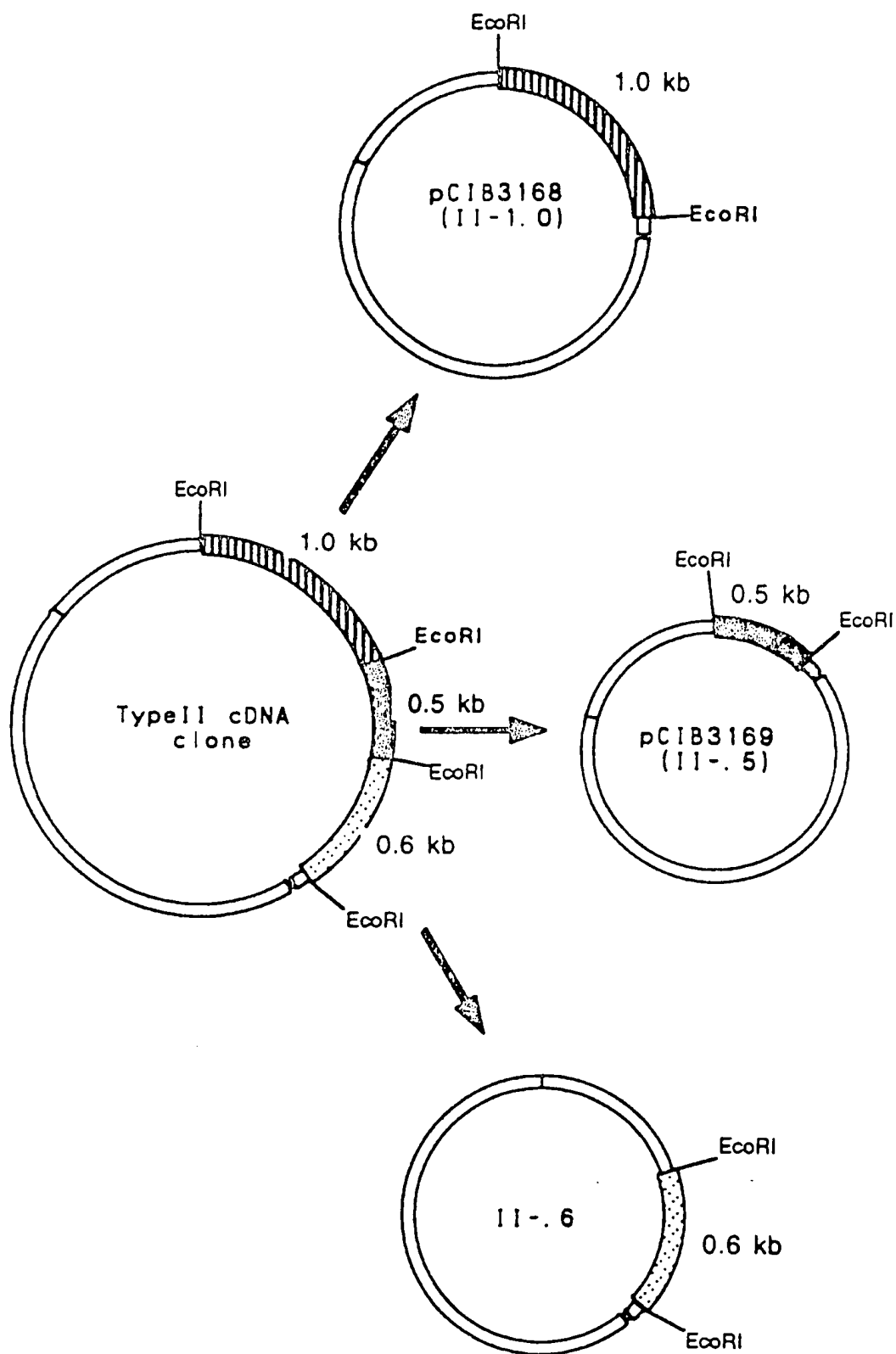
FIG. 30

Maize Pollen CDPK cDNA sequence

sequence contained in clones pCIB3168 and pCIB3169

```
  1 TG CAG ATC ATG CAC CAC CTC TCC GGC CAG CCC AAC GTG GTG GGC CTC CGC GGC GCG
    1▶Gln Ile Met His His Leu Ser Gly Gln Pro Asn Val Val Gly Leu Arg Gly Ala

 57 TAC GAG GAC AAG CAG AGC GTG CAC CTC GTC ATG GAG CTG TGC GCG GGC GGG GAG TTC
   19▶Tyr Glu Asp Lys Gln Ser Val His Leu Val Met Glu Leu Cys Ala Gly Gly Glu Leu

                    AvaI
114 TTC GAC CGC ATC ATC GCC CGG GGC CAG TAC ACG GAG CGC GGC GCC GCG GAG CTG CTG
   38▶Phe Asp Arg Ile Ile Ala Arg Gly Gln Tyr Thr Glu Arg Gly Ala Ala Glu Leu Leu

171 CGC GCC ATC GTG CAG ATC GTG CAC ACC TGC CAC TCC ATG GGG GTG ATG CAC CGG GAC
   57▶Arg Ala Ile Val Gln Ile Val His Thr Cys His Ser Met Gly Val Met His Arg Asp

         AvaI
228 ATC AAG CCC GAG AAC TTC CTG CTG CTC AGC AAG GAC GAG GAC GCG CCG CTC AAG GCC
   76▶Ile Lys Pro Glu Asn Phe Leu Leu Leu Ser Lys Asp Glu Asp Ala Pro Leu Lys Ala

285 ACC GAC TTC GGC CTC TCC GTC TTC TTC AAG GAG GGC GAG CTG CTC AGG GAC ATC GTC
   95▶Thr Asp Phe Gly Leu Ser Val Phe Phe Lys Glu Gly Glu Leu Leu Arg Asp Ile Val

                    AvaI
342 GGC AGC GCC TAC TAC ATC GCG CCC GAG GTG CTC AAG AGG AAG TAC GGC CCG GAG GCC
  114▶Gly Ser Ala Tyr Tyr Ile Ala Pro Glu Val Leu Lys Arg Lys Tyr Gly Pro Glu Ala

399 GAC ATC TGG AGC GTC GGC GTC ATG CTC TAC ATC TTC CTC GCC GGC GTG CCT CCC TTC
  133▶Asp Ile Trp Ser Val Gly Val Met Leu Tyr Ile Phe Leu Ala Gly Val Pro Pro Phe

456 TGG GCA GAG AAC GAG AAC GGC ATC TTC ACC GCC ATC CTG CGA GGG CAG CTT GAC CTC
  152▶Trp Ala Glu Asn Glu Asn Gly Ile Phe Thr Ala Ile Leu Arg Gly Gln Leu Asp Leu

513 TCC AGC GAG CCA TGG CCA CAC ATC TCG CCG GGA GCC AAG GAT CTC GTC AAG AAG ATG
  171▶Ser Ser Glu Pro Trp Pro His Ile Ser Pro Gly Ala Lys Asp Leu Val Lys Lys Met

570 CTC AAC ATC AAC CCC AAG GAG CGG CTC ACG GCG TTC CAG GTC CTC AAT CAC CCA TGG
  190▶Leu Asn Ile Asn Pro Lys Glu Arg Leu Thr Ala Phe Gln Val Leu Asn His Pro Trp

627 ATC AAA GAA GAC GGA GAC GCG CCT GAC ACG CCG CTT GAC AAC GTT GTT CTC GAC AGG
  209▶Ile Lys Glu Asp Gly Asp Ala Pro Asp Thr Pro Leu Asp Asn Val Val Leu Asp Arg

684 CTC AAG CAG TTC AGG GCC ATG AAC CAG TTC AAG AAA GCA GCA TTG AGG ATC ATA GCT
  228▶Leu Lys Gln Phe Arg Ala Met Asn Gln Phe Lys Lys Ala Ala Leu Arg Ile Ile Ala

741 GGG TGC CTA TCC GAA GAG GAG ATC ACA GGG CTG AAG GAG ATG TTC AAG AAC ATT GAC
  247▶Gly Cys Leu Ser Glu Glu Glu Ile Thr Gly Leu Lys Glu Met Phe Lys Asn Ile Asp

798 AAG GAT AAC AGC GGG ACC ATT ACC CTC GAC GAG CTC AAA CAC GGG TTG GCA AAG CAC
  266▶Lys Asp Asn Ser Gly Thr Ile Thr Leu Asp Glu Leu Lys His Gly Leu Ala Lys His

855 GGG CCC AAG CTG TCA GAC AGC GAA ATG GAG AAA CTA ATG GAA GCA GCT GAC GCT GAC
  285▶Gly Pro Lys Leu Ser Asp Ser Glu Met Glu Lys Leu Met Glu Ala Ala Asp Ala Asp

                    EcoRI
912 GGC AAC GGG TTA ATT GAC TAC GAC GAA TTC GTC ACC GCA ACA GTG CAT ATG AAC AAA
  304▶Gly Asn Gly Leu Ile Asp Tyr Asp Glu Phe Val Thr Ala Thr Val His Met Asn Lys
```

FIG. 30 (CONT)

```
 969 CTG GAT AGA GAA GAG CAC CTT TAC ACA GCA TTC CAG TAT TTC GAC AAG GAC AAC AGC
 323▶Leu Asp Arg Glu Glu His Leu Tyr Thr Ala Phe Gln Tyr Phe Asp Lys Asp Asn Ser

1026 GGG TAC ATT ACT AAA GAA GAG CTT GAG CAC GCC TTG AAG GAG CAA GGG TTG TAT GAC
 342▶Gly Tyr Ile Thr Lys Glu Glu Leu Glu His Ala Leu Lys Glu Gln Gly Leu Tyr Asp

1083 GCC GAT AAA ATC AAA GAC ATC ATC TCC GAT GCC GAC TCT GAC AAT GAT GGA AGG ATA
 361▶Ala Asp Lys Ile Lys Asp Ile Ile Ser Asp Ala Asp Ser Asp Asn Asp Gly Arg Ile

1140 GAT TAT TCA GAG TTT GTG GCG ATG ATG AGG AAA GGG ACG GCT GGT GCC GAG CCA ATG
 380▶Asp Tyr Ser Glu Phe Val Ala Met Met Arg Lys Gly Thr Ala Gly Ala Glu Pro Met

1197 AAC ATC AAG AAG AGG CGA GAC ATA GTC CTA TAG TGAAGTGAAGCAGCAAGTGTGTAATGTAATGTG
 399▶Asn Ile Lys Lys Arg Arg Asp Ile Val Leu • • •

1263 TATAGCAGCTCAAACAAGCAAATTTGTACATCTGTACACAAATGCAATGGGGTTACTTTTTGCAAAAAAAAAAAAAAAA
```

1340 AAAAAAAAAA

FIG. 31

81

inner leaf sheath

inner leaf whorl

green leaf

anther

pollen

silk

kernel

pith

root

FIG. 32

Lipman-Pearson Protein Alignment
Gap Penalty: 2, Gap Length Penalty: 12

Seq1 pol CDPK ptn	Seq2 rat pk2 ptn	Similarity Index	Gap Number	Gap Length	Consensus Length
1>551	1>528	36.5	4	4	297

```
pol CDPK  ptn  YSMGKELGRGQFGVTHLCTHRTSGEKLACKTIAKRKLAAREDVDDVRREVQIMHHLSGQPNVVGLRGAYE 162
               Y:  .ELG: G. F: V. :  C.. : TS. :   A K. I   : KL: AR:: :.: RE. : I  : L.  : PN: V L:::
rat pk2 ptn  YQLFEELGKGAFSVVRRCVKKTSTCEYAAKIINTKKLSARDH-QKLEREARICRLLK-HPNIVRLHDSIS 81


pol CDPK  ptn  CKQSVHLVMELCAGGELFDRIIARGQYTERGAAELLRAIVQIVHTCHSMGVMHRDIKPENFLLLSKDEDA 232
               :..    .LV.:L : GGELF:  I: AR. Y: E : A::  ::. I::  V:    H  ::: HRD: KPEN: LL  SK .: A
rat pk2 ptn  EEGFHYLVFDLVTGGELFEDIVAREYYSEADASHCIHQILESVNHIHQHDIVHRDLKPENLLLASKCKGA 15·


pol CDPK  ptn  PLKATDFGLSVFFKEGELLR-DIVGSAYYIAPEVLKRK-YGPEADIWSVGVMLYIFLAGVPPFWAENEHG 300
               :: K : DFGL::   :..:   :  ::. G:: Y: : PEVL::.  YG  . DIW:  GV: LYI: L. G  PPFW. E:::
rat pk2 ptn  AVKLADFGLAIEVQGEQQAWFGFAGTPGYLSPEVLRKCPYGKPVDIWACGVILYILLVGYPPFWDEDQHK 22·


pol CDPK  ptn  IFTAILRGQLDLSSEPWPHISPGAKDLVKKMLNINPKERLTAFQVLNHPWIKEDGDAPDTPLDNVVLDRL 370
               :: ·.I   G. D:: S   W  :: P. AK: L::: ML iNP   R: TA Q L: HPW: :  : ..:.    :  :: L
rat pk2 ptn  LYQQIKAGAYDFPSPEWDTVTPEAKNLINQMLTINPAKRITADCALKHPWVCCRSTVASMHHRQETVECL 29·


pol CDPK  ptn  KGFRAMNGFKKAALRII 387
               :: F. A. .:: K A L ..:
rat pk2 ptn  RKFNARRKLKGAIL⁻⁻M 308
```

82/94

FIG. 33

Lipman-Pearson Protein Alignment
Gap Penalty: 2, Gap Length Penalty: 12

Seq1 pol CDPK ptn	Seq2 humcama ptn	Similarity Index	Gap Number	Gap Length	Consensus Length
1>551	1>150	40.3	2	2	142

```
pol CDPK  ptn  LSEEEITGLKEMFKNIDKDNSGTITLDELKHGLAKHGPKLSDSEMEKLMEAADADGNGLIDYDEFVTATV 46C
               L:EE: I:  :KE F.  :DKD..GTIT .EL    :    G :  :::E::  :::.-DADGNG ID: EF:T  .
humcama ptn    LTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTIDFPEFLTMMA 74


pol CDPK  ptn  H-MNKLDREEHLYTAFQYFDKDNSGYITKEELEHALKECG-LYDADKIKDIISDADSDNDGRIDYSEFVA 528
               : M:. D.EE::  .AF: FDKD.:GYI: .EL H.:.: G    ..:.:.:: I.: AD D. DG:::Y. EFV.
humcama ptn    RKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEEVDEMIREADIDGDGQVNYEEFVQ 144


pol CDPK  ptn  MM 530
               MM
humcama ptn    MM 146
```

FIG. 34

Lipman-Pearson Protein Alignment
Gap Penalty: 2, Gap Length Penalty: 12

Seq1 pol CDPK ptn	Seq2 soybean CDPK ptn	Similarity Index	Gap Number	Gap Length	Consensus Length
1>551	1>509	62.4	1	1	464

```
pol CDPK  ptn    VLGRPMEDVRATYSMGKELGRGQFGVTHLCTHRTSGEKLACKTIAKRKLAAREDVDDVRREVQIMHHLSG  150
                 ||::::   :|::|::|: ||:||||:|:. ||:|:||: | |||:|:|||||::: ||  :||:||:::|||||:
soybean CDPK ptn VLPQRTQNIREVYEVGRKLGQGQFGTTFECTRRASGGKFACKSIPKRKLLCKEDYEDVWREIQIMHHLSE  91

pol CDPK  ptn    QPNVVGLRGAYEDKQSVHLVMELCAGGELFDRIIARGQYTERGAAELLRAIVQIVHTCHSMGVMHRDIKP  220
                 :: |||: :|: |||:.:|||||||: |||||||||: ::|: |: ||: ||:|  |. | : ||| ||||||  ||
soybean.CDPK ptn HANVVRIEGTYEDSTAVHLVMELCEGGELFDRIVQKGHYSERQAARLIKTIVEVVEACHSLGVMHRDLKP  161

pol CDPK  ptn    ENFLLLSKDEDAPLKATDFGLSVFFKEGELLRDIVGSAYYIAPEVLKRKYGPEADIWSVGVMLYIFLAGV  290
                 ||||  .::|||||:|||||||||||||||||:|,|| :|  |||:|| ||||||::. ||||:| ||:||: |||:|:||
soybean CDPK ptn ENFLFDTIDEDAKLKATDFGLSVFYKPGESFCDVVGSPYYVAPEVLRKLYGPESDVWSAGVILYILLSGV  231

pol CDPK  ptn    PPFWAENENGIFTAILRGQLDLSSEPWPHISPGAKDLVKKMLNINPKERLTAFQVLNHPWIKEDGDAPDT  360
                 |||||||:|:|||  :|| |:|| |:||  :|||||:||::  ||||::  |||:  ||||:||||:.|| ||||: ':.|||:
soybean CDPK ptn PPFWAESEPGIFRQILLGKLDFHSEPWPSISDSAKDLIRKMLDQNPKTRLTAHEVLRHPWIVDDNIAPDK  301

po  CDPK  ptn    PLDNVVLDRLKQFRAMNQFKKAALRIIAGCLSEEEITGLKEMFKNIDKDNSGTITLDELKHGLAKHGPKL  430
                 |||::.|| |||||| |||::.||:||| ||::.|||||||| |||| || || ||||||||| ||||:.||:: |::|
soybean CDPK ptn PLDSAVLSRLKQFSAMNKLKKMALRVIAERLSEEEIGGLKELFKMIDTDNSGTITFDELKDGLKRVGSEL  371

pol CDPK  ptn    SDSEMEKLMEAADADGNGLIDYDEFVTATVHMNKLDREEHLYTAFQYFDKDNSGYITKEELEHALKEQGL  500
                 ::||: :|| |||:|::|,|||:||:::|||| |||:|||  |:||.|||||:|||||. |.:|| |:||
soybean CDPK ptn MESEIKDLMDAADIDKSGTIDYGEFIAATVHLNKLEREENLVSAFSYFDKDGSGYITLDEIQCACKDFGL  441

po  CDPK  ptn    YDADKIKDIISDADSDNDGRIDYSEFVAMMRKGTAGAEPMNIKK  544
                 |:. |:':|:::|:|||:|||:|||:|::||:||:::|::::  ::|
soybean CDPK ptn -DDIHIDDMIKEIDDNDGDIDYGEFAAMMRKGNGGIGRRTMRK  484
```

FIG. 35

pol CDPK gene Map (1 > 4165)   Site and Sequence
Enzymes :   6 of 198 enzymes (Filtered)
Settings :   Circular, Certain Sites Only, Standard Genetic Code

```
TTAGTAACACCTCTCCAATCGCTTGGGTTGGCACATTCTTAGCTTTTATCACATTTTAAGAAATAGAGTTCACCACCTTC
                                                                                  80

AAAATATGCCTATACAATGAATGATGCTTGGATGCAATATAGCTAGATTCAACTAGCTATATATGGTCAATAGAACCCTG
                                                                                  160

TGAGCACCTCACAAACACGACTTCAATTTTGAGACCCTAAGCGAGTAAATGGTTAAAGTCCTCTTATTATTAGTCTTAGG
                                                                                  240

ACTTCTCCTTGCTAAATGCTTGTCAGCGATCTATATATCTTCCCCACTGCGGGAGATACTATATATAGGGCCTTGGACCT
                                                                                  320

CTAGGGTATCTCAAAGGCCTAGTCACAACAATTCTCAACAGTATTTAATTTTATACATGTATGAACAGTGTAGGAATTTG
                                                                                  400

AGTGCCCAACCCAAGAGTGGGAGGTGTAAATTGGGTAGCTAAACTTAAATAGGGCTCTTCTTATTTAGGTTTATCTAGTC
                                                                                  480

TCTACTTAGACTAATTCAGAAAGAATTTTACAACCTATGGTTAATCATATCTCTAGTCTAAGCAAATTTAGGAAAGTTAA
                                                                                  560

AAGCACACAATTAGGCACATGTGAAAGATGTGTATGGTAAGTAAAAGACTTATAAGGAAAAAGTGGGTGAATCCTCAAGA
                                                                                  640

TGTGGTGGTATATCCCAATGATATTAGATGCCAGAATATAGGGGGGGAAATCGATGTATACCATCTCTACCAGGATACCTG
                                                                                  720

TGCGGACTGTGCAACTGACACATGGACCATGGTGTCTTCTTAGATTTGGTTATTAGCTAATTGCGCTACAACTTGTTCAA
                                                                                  800

GGCTAGACCAAATTAAAAAAACTAATATTAAACATAAAAAGTTAGGCAAACTATAGTAAATTATGCAGCGATCCAACAACA
                                                                                  880

AGCCATGTCTCGTGGGTCATGAGCCACGCGTCGGCCATACACCCACATGATGTTTCCATACGGATGGTCCTTATGCAATT
                                                                                  960

TTGTCTGCAAAACACAAGCCTTAATACAGCCACGCGACAATCATGGAAGTGGTCGTTTTAGGTCCTCATCATGAAGTTCA
                                                                                  1040

GGGAAAACGCATCAAATGTAATGCAGAGAAATGGTATTTCTTCTCTTGTAAATCAGGGAGAGGAGTACCATCAGTACAGA
                                                                                  1120
```

         EcoR I

```
TTCAGAATCAGAATTCAGTCTTCCAACGACAATAATCGCAGCATCTTGTAAAAATTTGCAGAAACTTCTGTTTGACTTGT
                                                                                  1200

AGCCCTGACCTTTGCAAATATTTGAAGTTGTGCCTGCTGACACAACTTCAATCTGGAAGTGCTGTTGATCAGTTTTGCCA
                                                                                  1280

GAAACAGCAAGCAGCCTATATATATCTGTCACGAGACACCCTGCCGCCCTCTTCTTTCCCGCCATTCCCTCCCTACCCTT
                                                                                  1360
```

FIG. 35 (CONT)

pol CDPK gene Map (1 > 4165)   Site and Sequence

Xba I

CAAAATCTAGAAACCTTTTTTTTTCCTCCCGATACGCCCCTCCATCTCTCGCCGTTCATGTCCGTGGCTGGCTGCCCTCC ┤ 1440

⌐ mRNA start ⌐

GTGGGAGCAGGCGGCCGCACTCGTTCCCCGCCGCAGCCATGGGCCAGTGCTGCTCCAAGGGCGCCGGAGAGGCCCCGCCA ┤ 1520

═══════════════════ EXON 1 ═══════════════════

CCGAGGCGCCAAACGGCAGGCGCCAAGCCGCGGGCGTCCGCGAACAACGCCGACGGACAACGGGCGTCGTCCTCGTCCGC ┤ 1600

═══════════════════ EXON 1 ═══════════════════

GGTGGCTGCTGCCGCTGCTGCTGCCGGTGGTGGTGGCGGCGGCACGACGAAGCCGGCCTCACCCACCGGCGGCGCCAGGG ┤ 1680

═══════════════════ EXON 1 ═══════════════════

CCAGCTCCGGCAGCAAACCGGCGGCGGCCGTGGGCACGGTGCTGGGCCGGCCCATGGAGGACGTGCGCGCGACCTACTCG ┤ 1760

═══════════════════ EXON 1 ═══════════════════

Ava I

ATGGGCAAGGAGCTCGGGCGCGGGCAGTTCGGCGTGACGCACCTGTGCACGCACCGGACGAGCGGCGAGAAGCTGGCGTG ┤ 1840

═══════════════════ EXON 1 ═══════════════════

CAAGACGATCGCGAAGCGGAAGCTGGCGGCCAGGGAGGACGTGGACGACGTGCGGCGGGAGGTGCAGATCATGCACCACC ┤ 1920

═══════════════════ EXON 1 ═══════════════════

TCTCCGGCCAGCCCAACGTGGTGGGCCTCCGCGGCGCGTACGAGGACAAGCAGAGCGTGCACCTCGTCATGGAGCTGTGC ┤ 2000

═══════════════════ EXON 1 ═══════════════════

Ava I

GCGGGCGGGGAGCTCTTCGACCGCATCATCGCCCGGGGGCCAGTACACGGAGCGCGGCGCCGCGGAGCTGCTGCGCGCCAT ┤ 2080

═══════════════════ EXON 1 ═══════════════════

FIG. 35 (CONT)

pol CDPK gene Map (1 > 4165)   Site and Sequence

Ava I

CGTGCAGATCGTGCACACCTGCCACTCCATGGGGGTGATGCACCGGGACATCAAGCCCGAGAACTTCCTGCTGCTCAGCA   2160

———————————————————————— EXON 1 ————————————————————————

AGGACGAGGACGCGCCGCTCAAGGCCACCGACTTCGGCCTCTCCGTCTTCTTCAAGGAGGGCGAGCTGCTCAGGGACATC   2240

———————————————————————— EXON 1 ————————————————————————

Ava I

GTCGGCAGCGCCTACTACATCGCGCCCGAGGTGCTCAAGAGGAAGTACGGCCCGGAGGCCGACATCTGGAGCGTCGGCGT   2320

———————————————————————— EXON 1 ————————————————————————

Bam H I

CATGCTCTACATCTTCCTCGCCGGCGTGCCTCCCTTCTGGGCAGGTCGGATCCGTCCGTGTTCGTCCTAGACGATATACA   2400

———————————— EXON 1 ————————————   ———————————— INTRON 1 ————————————

GAACCCGACGATGGATTTGCTTCTCAGCCCTGTTCTTGCATCACCAGAGAACGAGAACGGCATCTTCACCGCCATCCTGC   2480

———————————— INTRON 1 ————————————   ———————————— EXON 2 ————————————

GAGGGCAGCTTGACCTCTCCAGCGAGCCATGGCCACACATCTCGCCGGGAGCCAAGGATCTCGTCAAGAAGATGCTCAAC   2560

———————————————————————— EXON 2 ————————————————————————

ATCAACCCCAAGGAGCGGCTCACGGCGTTCCAGGTCCTCAGTAAGTACCCAGATCGTTGCTGTCATACACTCATATGAAT   2640

———————————— EXON 2 ————————————   ———————————— INTRON 2 ————————————

TGTATCGTTCATGAGCAACGATCGAGCGGATTTGGTGAACTTGTAGATCACCCATGGATCAAAGAAGACGGAGACGCGCC   2720

———————————— INTRON 2 ————————————   ———————————— EXON 3 ————————————

TGACACGCCGCTTGACAACGTTGTTCTCGACAGGCTCAAGCAGTTCAGGGCCATGAACCAGTTCAAGAAAGCAGCATTGA   2800

———————————————————————— EXON 3 ————————————————————————

FIG. 35 (CONT)

pol CDPK gene Map (1 > 4165)   Site and Sequence

GGGTACATTATCTGATAAAAGCTCCACAAATACAACTTCTGAAGAACAGCAATGCTTACACGGCAGAATTTTCATTATAA
                                                                                                      2880

———————————— INTRON 3 ————————————

ATGCTCTTGATGACATAATGTTAGATCATAGCTGGGTGCCTATCCGAAGAGGAGATCACAGGGCTGAAGGAGATGTTCAA
                                                                                                      2960

———— INTRON 3 ————   ———————— EXON 4 ————————

GAACATTGACAAGGATAACAGCGGGACCATTACCCTCGACGAGCTCAAACACGGGTTGGCAAAGCACGGGCCCAAGCTGT
                                                                                                      3040

———————————— EXON 4 ————————————

CAGACAGCGAAATGGAGAAACTAATGGAAGCAGTGAGTTTTCAGAGTACAATCTTAAAAAAAGGAATTGTGATTCTTTTC
                                                                                                      3120

———— EXON 4 ————   ———————— INTRON 4 ————————

AAAATGAAGAAGTAATCTGAAAACATCCCTGCTGAAATGCTTTATACATTTCCAGGCTGACGCTGACGGCAACGGGTTAA
                                                                                                      3200

———————— INTRON 4 ————————   ———— EXON 5 ————

EcoR I

TTGACTACGACGAATTCGTCACCGCAACAGTGCATATGAACAAACTGGATAGAGAAGAGCACCTTTACACAGCATTCCAG
                                                                                                      3280

———————————— EXON 5 ————————————

                                                                              EcoR I

TATTTCGACAAGGACAACAGCGGGTAAGTTGAACGTTAAAATGATACAGCTGGTACCTGAATTCTGGACAACACATATCA
                                                                                                      3360

———— EXON 5 ————   ———————— INTRON 5 ————————

TAACAGGACACATATATAATTCGTTTATCTCACAGGTACATTACTAAAGAAGAGCTTGAGCACGCCTTGAAGGAGCAAGG
                                                                                                      3440

———————— INTRON 5 ————————   ———— EXON 6 ————

GTTGTATGACGCCGATAAAATCAAAGACATCATCTCCGATGCCGACTCTGACAATGTAAGGAACAAACATTATTTAAATT
                                                                                                      3520

———————————— EXON 6 ————————————   ———— INTRON 6 ————

FIG. 35 (CONT)

pol CDPK gene Map (1 > 4165)   Site and Sequence

TCAGCCGACAAACTAAACTATAGAAACCACATCATGATATCAAATTTTGAGGTGGCGGTGCTACAGAAATAGAACCCAGT
     3600

———————————— INTRON 6 ————————————

ACACCAAAATGACTAACTTGTCATGATTAGTTGTTCCTCGTAACTGAACATTTGTGTTCTTAGTTTCTTATTGTTAAACC
     3680

———————————— INTRON 6 ————————————

AAAGACTTAAATTCACTTTTGCACATGCAGGATGGAAGGATAGATTATTCAGAGTTTGTGGCGATGATGAGGAAAGGGAC
     3760

——— INTRON 6 ———   ——— EXON 7 ———

GGCTGGTGCCGAGCCAATGAACATCAAGAAGAGGCGAGACATAGTCCTATAGTGAAGTGAAGCAGWAAGTGTGTAATGTA
     3840

——————— EXON 7 ———————

ATGTGTATAGCAGCTCAAACAAGCAAATTTGTACATCTGTACACAAATGCAATGGGGTTACTTTTGCAACTTAGTTCATG
     3920

GATGGTTGTGTACGTTGTGCTATTGATTGCAAGTGATTTGAAAGACATGCATACTTAGGAACTGAGAAAGATAGATCTAC
     4000

TACTGCTAGAGACAGAACAATAGGATKKYAATTCAGYAAGTGYGTATTTCAGAAGACTACAGCTGGCATCTATTATTCTC
     4080

ATTGTCCTCGCAAAAATACTGATGATGCATTTGAGAGAACAATATGCAACAAGATCGAGCTCCCTATAGTGAGTCGTATT
   4160

AGGCC
———→ 4165

FIG. 36

Fig. 37
E:\GF2-PAT.SEQ 9/21/92

NAME: pCIB5515   7415 BPS DNA   CIRCULAR   UPDATED   9/21/92
DESCRIPTION: pGF1 with synthetic geiser fix.

* * *   S E Q U E N C E   * * *


```
    1  ATGGACAACA ACCCCAACAT CAACGAGTGC ATCCCCTACA ACTGCCTGAG CAACCCCGAG
       MetAspAsn AsnProAsn IleAsnGluCys IleProTyr AsnCysLeu SerAsnProGlu

   61  GTGGAGGTGC TGGGCGGCGA GCGCATCGAG ACCGGCTACA CCCCCATCGA CATCAGCCTG
       ValGluVal LeuGlyGly GluArgIleGlu ThrGlyTyr ThrProIle AspIleSerLeu

  121  AGCCTGACCC AGTTCCTGCT GAGCGAGTTC GTGCCCGGCG CCGGCTTCGT GCTGGGCCTG
       SerLeuThr GlnPheLeu LeuSerGluPhe ValProGly AlaGlyPhe ValLeuGlyLeu

  181  GTGGACATCA TCTGGGGCAT CTTCGGCCCC AGCCAGTGGG ACGCCTTCCT GGTGCAGATC
       ValAspIle IleTrpGly IlePheGlyPro SerGlnTrp AspAlaPhe LeuValGlnIle

  241  GAGCAGCTGA TCAACCAGCG CATCGAGGAG TTCGCCCGCA ACCAGGCCAT CAGCCGCCTG
       GluGlnLeu IleAsnGln ArgIleGluGlu PheAlaArg AsnGlnAla IleSerArgLeu

  301  GAGGGCCTGA GCAACCTGTA CCAAATCTAC GCCGAGAGCT TCCGCGAGTG GGAGGCCGAC
       GluGlyLeu SerAsnLeu TyrGlnIleTyr AlaGluSer PheArgGlu TrpGluAlaAsp

  361  CCCACCAACC CCGCCCTGCG CGAGGAGATG CGCATCCAGT TCAACGACAT GAACAGCGCC
       ProThrAsn ProAlaLeu ArgGluGluMet ArgIleGln PheAsnAsp MetAsnSerAla

  421  CTGACCACCG CCATCCCCCT GTTCGCCGTG CAGAACTACC AGGTGCCCCT GCTGAGCGTG
       LeuThrThr AlaIlePro LeuPheAlaVal GlnAsnTyr GlnValPro LeuLeuSerVal

  481  TACGTGCAGG CCGCCAACCT GCACCTGAGC GTGCTGCGCG ACGTCAGCGT GTTCGGCCAG
       TyrValGln AlaAlaAsn LeuHisLeuSer ValLeuArg AspValSer ValPheGlyGln

  541  CGCTGGGGCT TCGACGCCGC CACCATCAAC AGCCGCTACA ACGACCTGAC CCGCCTGATC
       ArgTrpGly PheAspAla AlaThrIleAsn SerArgTyr AsnAspLeu ThrArgLeuIle

  601  GGCAACTACA CCGACCACGC CGTGCGCTGG TACAACACCG GCCTGGAGCG CGTGTGGGGT
       GlyAsnTyr ThrAspHis AlaValArgTrp TyrAsnThr GlyLeuGlu ArgValTrpGly

  661  CCCGACAGCC GCGACTGGAT CAGGTACAAC CAGTTCCGCC GCGAGCTGAC CCTGACCGTG
       ProAspSer ArgAspTrp IleArgTyrAsn GlnPheArg ArgGluLeu ThrLeuThrVal

  721  CTGGACATCG TGAGCCTGTT CCCCAACTAC GACAGCCGCA CCTACCCCAT CCGCACCGTG
       LeuAspIle ValSerLeu PheProAsnTyr AspSerArg ThrTyrPro IleArgThrVal

  781  AGCCAGCTGA CCCGCGAGAT TTACACCAAC CCCGTGCTGG AGAACTTCGA CGGCAGCTTC
       SerGlnLeu ThrArgGlu IleTyrThrAsn ProValLeu GluAsnPhe AspGlySerPhe

  841  CGCGGCAGCG CCCAGGGCAT CGAGGGCAGC ATCCGCAGCC CCCACCTGAT GGACATCCTG
       ArgGlySer AlaGlnGly IleGluGlySer IleArgSer ProHisLeu MetAspIleLeu

  901  AACAGCATCA CCATCTACAC CGACGCCCAC CGCGGCGAGT ACTACTGGAG CGGCCACCAG
       AsnSerIle ThrIleTyr ThrAspAlaHis ArgGlyGlu TyrTyrTrp SerGlyHisGln

  961  ATCATGGCCA GCCCCGTCGG CTTCAGCGGC CCCGAGTTCA CCTTCCCCCT GTACGGCACC
       IleMetAla SerProVal GlyPheSerGly ProGluPhe ThrPhePro LeuTyrGlyThr

 1021  ATGGGCAACG CTGCACCTCA GCAGCGCATC GTGGCACAGC TGGGCCAGGG AGTGTACCGC
       MetGlyAsn AlaAlaPro GlnGlnArgIle ValAlaGln LeuGlyGln GlyValTyrArg

 1081  ACCCTGAGCA GCACCCTGTA CCGTCGACCT TTCAACATCG GCATCAACAA CCAGCAGCTG
       ThrLeuSer SerThrLeu TyrArgArgPro PheAsnIle GlyIleAsn AsnGlnGlnLeu

 1141  AGCGTGCTGG ACGGCACCGA GTTCGCCTAC GGCACCAGCA GCAACCTGCC CAGCGCCGTG
       SerValLeu AspGlyThr GluPheAlaTyr GlyThrSer SerAsnLeu ProSerAlaVal

 1201  TACCGCAAGA GCGGCACCGT GGACAGCCTG GACGAGATCC CCCCTCAGAA CAACAACGTG
       TyrArgLys SerGlyThr ValAspSerLeu AspGluIle ProProGln AsnAsnAsnVal
```

```
1261  CCACCTCGAC AGGGCTTCAG CCACCGTCTG AGCCACGTGA GCATGTTCCG CAGTGGCTTC
      ProProArg  GlnGlyPhe  SerHisArgLeu SerHisVal SerMetPhe  ArgSerGlyPhe

1321  AGCAACAGCA GCGTGAGCAT CATCCGTGCA CCTATGTTCA GCTGGATTCA CCGCAGTGCC
      SerAsnSer  SerValSer  IleIleArgAla ProMetPhe SerTrpIle  HisArgSerAla

1381  GAGTTCAACA ACATCATCCC CAGCAGCCAG ATCACCCAGA TCCCCCTGAC CAAGAGCACC
      GluPheAsn  AsnIleIle  ProSerSerGln IleThrGln IleProLeu  ThrLysSerThr

1441  AACCTGGGCA GCGGCACCAG CGTGGTGAAG GGCCCCGGCT TCACCGGCGG CGACATCCTG
      AsnLeuGly  SerGlyThr  SerValValLys GlyProGly PheThrGly  GlyAspIleLeu

1501  CGCCGCACCA GCCCCGGCCA GATCAGCACC CTGCGCGTGA ACATCACCGC CCCCCTGAGC
      ArgArgThr  SerProGly  GlnIleSerThr LeuArgVal AsnIleThr  AlaProLeuSer

1561  CAGCGCTACC GCGTCCGCAT CCGCTACGCC AGCACCACCA ACCTGCAGTT CCACACCAGC
      GlnArgTyr  ArgValArg  IleArgTyrAla SerThrThr AsnLeuGln  PheHisThrSer

1621  ATCGACGGCC GCCCCATCAA CCAGGGCAAC TTCAGCGCCA CCATGAGCAG CGGCAGCAAC
      IleAspGly  ArgProIle  AsnGlnGlyAsn PheSerAla ThrMetSer  SerGlySerAsn

1681  CTGCAGAGCG GCAGCTTCCG CACCGTGGGC TTCACCACCC CCTTCAACTT CAGCAACGGC
      LeuGlnSer  GlySerPhe  ArgThrValGly PheThrThr ProPheAsn  PheSerAsnGly

1741  AGCAGCGTGT CCACCCTGAG CGCCCACGTG TTCAACAGCG GCAACGAGGT GTACATCGAC
      SerSerVal  PheThrLeu  SerAlaHisVal PheAsnSer GlyAsnGlu  ValTyrIleAsp

1801  CGCATCGAGT TCGTGCCCGC CGAGGTGACC TTCGAGGCCA GTACGACCT GGAGAGGGCT
      ArgIleGlu  PheValPro  AlaGluValThr PheGluAla GluTyrAsp  LeuGluArgAla

1861  CAGAAGGCCG TGAACGAGCT GTTCACCAGC AGCAACCAGA TCGGCCTGAA GACCGACGTG
      GlnLysAla  ValAsnGlu  LeuPheThrSer SerAsnGln IleGlyLeu  LysThrAspVal

1921  ACCGACTACC ACATCGATCA AGTATCCAAT TTAGTTGAGT GTTTATCTGA TGAATTTTGT
      ThrAspTyr  HisIleAsp  GlnValSerAsn LeuValGlu CysLeuSer  AspGluPheCys

1981  CTGGATGAAA AAAAAGAATT GTCCGAGAAA GTCAAACATG CGAAGCGACT TAGTGATGAG
      LeuAspGlu  LysLysGlu  LeuSerGluLys ValLysHis AlaLysArg  LeuSerAspGlu

2041  CGGAATTTAC TTCAAGATCC AAACTTTAGA GGGATCAATA GACAACTAGA CCGTGGCTGG
      ArgAsnLeu  LeuGlnAsp  ProAsnPheArg GlyIleAsn ArgGlnLeu  AspArgGlyTrp

2101  AGAGGAAGTA CGGATATTAC CATCCAAGGA GGCGATGACG TATTCAAAGA GAATTACGTT
      ArgGlySer  ThrAspIle  ThrIleGlnGly GlyAspAsp ValPheLys  GluAsnTyrVal

2161  ACGCTATTGG GTACCTTTGA TGAGTGCTAT CCAACGTATT TATATCAAAA AATAGATGAG
      ThrLeuLeu  GlyThrPhe  AspGluCysTyr ProThrTyr LeuTyrGln  LysIleAspGlu

2221  TCGAAATTAA AAGCCTATAC CCGTTACCAA TTAAGAGGGT ATATCGAAGA TAGTCAAGAC
      SerLysLeu  LysAlaTyr  ThrArgTyrGln LeuArgGly TyrIleGlu  AspSerGlnAsp

2281  TTAGAAATCT ATTTAATTCG CTACAATGCC AAACACGAAA CAGTAAATGT GCCAGGTACG
      LeuGluIle  TyrLeuIle  ArgTyrAsnAla LysHisGlu ThrValAsn  ValProGlyThr

2341  GGTTCCTTAT GGCCGCTTTC AGCCCCAAGT CCAATCGGAA AATGTGGGGA GCCGAATCGA
      GlySerLeu  TrpProLeu  SerAlaProSer ProIleGly LysCysGly  GluProAsnArg

2401  TGCGCTCCGC ACCTGGAGTG GAACCCGGAC CTAGACTGCA GCTGCAGGGA CGGGGAGAAG
      CysAlaPro  HisLeuGlu  TrpAsnProAsp LeuAspCys SerCysArg  AspGlyGluLys

2461  TGCGCCCATC ATTCCCATCA TTTCTCCTTG GACATTGATG TTGGATGTAC AGACTTAAAT
      CysAlaHis  HisSerHis  HisPheSerLeu AspIleAsp ValGlyCys  ThrAspLeuAsn

2521  GAGGACTTAG GTGTATGGGT GATATTCAAG ATTAAGACGC AAGATGGCCA TGCAAGACTA
      GluAspLeu  GlyValTrp  ValIlePheLys IleLysThr GlnAspGly  HisAlaArgLeu

2581  GGAAATCTAG AATTTCTCGA AGAGAAACCA TTAGTAGGAG AAGCACTAGC TCGTGTGAAA
      GlyAsnLeu  GluPheLeu  GluGluLysPro LeuValGly GluAlaLeu  AlaArgValLys
```

Fig. 37 (CONT)
E:\GF2-PAT.SEQ 9/21/92

```
2641  AGAGCGGAGA AAAAATGGAG AGACAAACGT GAAAAATTGG AATGGGAAAC AAATATTGTT
      ArgAlaGlu  LysLysTrp  ArgAspLysArg GluLysLeu  GluTrpGlu  ThrAsnIleVal

2701  TATAAAGAGG CAAAAGAATC TGTAGATGCT TTATTTGTAA ACTCTCAATA TGATAGATTA
      TyrLysGlu  AlaLysGlu  SerValAspAla LeuPheVal  AsnSerGln  TyrAspArgLeu

2761  CAAGCGGATA CCAACATCGC GATGATTCAT GCGGCAGATA AACGCGTTCA TAGCATTCGA
      GlnAlaAsp  ThrAsnIle  AlaMetIleHis AlaAlaAsp  LysArgVal  HisSerIleArg

2821  GAAGCTTATC TGCCTGAGCT GTCTGTGATT CCGGGTGTCA ATGCGGCTAT TTTTGAAGAA
      GluAlaTyr  LeuProGlu  LeuSerValIle ProGlyVal  AsnAlaAla  IlePheGluGlu

2881  TTAGAAGGGC GTATTTTCAC TGCATTCTCC CTATATGATG CGAGAAATGT CATTAAAAAT
      LeuGluGly  ArgIlePhe  ThrAlaPheSer LeuTyrAsp  AlaArgAsn  ValIleLysAsn

2941  GGTGATTTTA ATAATGGCTT ATCCTGCTGG AACGTGAAAG GGCATGTAGA TGTAGAAGAA
      GlyAspPhe  AsnAsnGly  LeuSerCysTrp AsnValLys  GlyHisVal  AspValGluGlu

3001  CAAAACAACC ACCGTTCGGT CCTTGTTGTT CCGGAATGGG AAGCAGAAGT GTCACAAGAA
      GlnAsnAsn  HisArgSer  ValLeuValVal ProGluTrp  GluAlaGlu  ValSerGlnGlu

3061  GTTCGTGTCT GTCCGGGTCG TGGCTATATC CTTCGTGTCA CAGCGTACAA GGAGGGATAT
      ValArgVal  CysProGly  ArgGlyTyrIle LeuArgVal  ThrAlaTyr  LysGluGlyTyr

3121  GGAGAAGGTT GCGTAACCAT TCATGAGATC GAGAACAATA CAGACGAACT GAAGTTTAGC
      GlyGluGly  CysValThr  IleHisGluIle GluAsnAsn  ThrAspGlu  LeuLysPheSer

3181  AACTGTGTAG AAGAGGAAGT ATATCCAAAC AACACGGTAA CGTGTAATGA TTATACTGCG
      AsnCysVal  GluGluGlu  ValTyrProAsn AsnThrVal  ThrCysAsn  AspTyrThrAla

3241  ACTCAAGAAG AATATGAGGG TACGTACACT TCTCGTAATC GAGGATATGA CGGAGCCTAT
      ThrGlnGlu  GluTyrGlu  GlyThrTyrThr SerArgAsn  ArgGlyTyr  AspGlyAlaTyr

3301  GAAAGCAATT CTTCTGTACC AGCTGATTAT GCATCAGCCT ATGAAGAAAA AGCATATACA
      GluSerAsn  SerSerVal  ProAlaAspTyr AlaSerAla  TyrGluGlu  LysAlaTyrThr

3361  GATGGACGAA GAGACAATCC TTGTGAATCT AACAGAGGAT ATGGGGATTA CACACCACTA
      AspGlyArg  ArgAspAsn  ProCysGluSer AsnArgGly  TyrGlyAsp  TyrThrProLeu

3421  CCAGCTGGCT ATGTGACAAA AGAATTAGAG TACTTCCCAG AAACCGATAA GGTATGGATT
      ProAlaGly  TyrValThr  LysGluLeuGlu TyrPhePro  GluThrAsp  LysValTrpIle

3481  GAGATCGGAG AAACGGAAGG AACATTCATC GTGGACAGCG TGGAATTACT TCTTATGGAG
      GluIleGly  GluThrGlu  GlyThrPheIle ValAspSer  ValGluLeu  LeuLeuMetGlu

3541  GAATAA
      Glu---
```

Fig. 38

pCIB5515
7415 bps

GF2-CryIA(b)

PEPC Intron I36 terminator

amp

lacZ''

35S-5'R

ApaI,1470
BstEII,1824
ClaI,1933
KpnI,2170
ClaI,2396
XbaI,2586
MluI,2802

SacI,3547
BglII,3553
ClaI,3640
3744,KpnI
3750,EcoRI

7406,BamHI
7264,EcoRV

6577,StuI
6409,XbaI
6391,SphI